PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A01K 67/027, 67/033, C12N 15/00 | A2

(11) International Publication Number: WO 98/28971

(43) International Publication Date: 9 July 1998 (09.07.98)

(21) International Application Number: PCT/US97/23819

(22) International Filing Date: 31 December 1997 (31.12.97)

(30) Priority Data:

08/778,645	3 January 1997 (03.01.97)	US
60/080,222	3 January 1997 (03.01.97)	US

(71) Applicant (*for all designated States except US*): UNIVERSITY TECHNOLOGY CORPORATION [US/US]; Suite 250, 3101 Iris Avenue, Boulder, CO 80301 (US).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): LINK, Christopher [US/US]; 1837 Spruce Street, Boulder, CO 80302 (US).

(74) Agents: MONROY, Gladys, H. et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: BETA-AMYLOID TOXICITY

(57) Abstract

The present invention relates to methods and compositions for visualization of the toxic effects of transgenes *in vivo*. In particular, the present invention provides methods and compositions for the production and use of transgenic, including dually transgenic, *Caenorhabditis elegans* for visualization of the toxic effects of β-amyloid accumulation *in vivo*.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

BETA-AMYLOID TOXICITY

FIELD OF THE INVENTION

The present invention is in the field of medicine and molecular biology. In particular, the invention provides transgenic non-human animals in which expression of a reporter gene is induced by a toxic transgene. The invention also provides methods and compositions for *in vivo* visualization of the toxicity associated with toxic proteins. In addition, the present invention provides methods and compositions for the production and use of transgenic non-human animals for screening candidate drugs to assess and treat toxicity.

BACKGROUND OF THE INVENTION

Investigation of the molecular mechanisms of various diseases has suggested that many pathologies are caused by expression of proteins that are immediately or gradually deleterious to the cells where they are expressed. Among these diseases are Alzheimer's disease, prion diseases, Huntington's disease, and amyotrophic lateral sclerosis. Several of these diseases are thought to result from aberrant folding of proteins, which results in the accumulation of toxic proteins or protein aggregates.

Alzheimer's disease is associated with the majority of dementia cases in the United States, with an estimated 2 million people afflicted with the disease, and a mortality rate of approximately 100,000 people per year (*See,* R.W.P. Cutler, "Degenerative and Hereditary Diseases," *in Medicine*, Scientific American, New York, (1988), pages 11 (IV):1-13; and R. Katzman (1986) *N. Engl. J. Med.* **314**:964). It has been estimated that the total cost for nursing home care alone of Alzheimer's patients exceeds $13 billion/year (*See,* M.M. Heckler (1985) *Am. Psychol.* **40**:1240). According to the Centers for Disease Control (CDC), mortality due to Alzheimer's disease in the United States increased 10-fold between 1979 and 1987 (0.4 per 100,000 to 4.2 per 100,000) (*See,* "Reported death rate for Alzheimer's is up tenfold since 1979," (1990) *Clin. Psychiatr. News* **18**:21).

Patients suffering from Alzheimer's disease typically suffer progressive memory deficit, progressive decline in cognitive functions, anxiety, depression, visuospatial and speech deficits, delusions, personality changes, motor skill deterioration, loss of verbal

ability, and incontinence. Eventually, patients are completely incapacitated and disoriented, requiring total care. The course of Alzheimer's disease ranges from less than three years to over 20 years before death occurs. However, in typical cases, it progresses at a fairly constant rate, with an average duration of 6 to 10 years.

Alzheimer's disease is not the only cause of dementia. Indeed, there are over fifty recognized causes of dementia. As some causes of dementia are amenable to treatment, differential diagnosis of patients suffering dementia is particularly important. Because at least 20% of clinically diagnosed patients were found at autopsy to have had conditions other than Alzheimer's disease, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and Alzheimer's Disease and Related Disorders Association (ADRDA) refined the clinical diagnostic criteria for Alzheimer's disease (*See*, M.A. Jenike, "Psychiatry," *in Medicine*, Scientific American, New York, [1991], pages 13 (V):1-5). Based on these criteria, the diagnosis of Alzheimer's disease may be "definite" (*i.e.*, requiring examination of brain tissue), "probable" (*i.e.*, patients have deficits in two or more areas of cognition, insidious onset of disease, progressive worsening of memory and other cognitive functions, and normal consciousness levels), or "possible" (*i.e.*, patents meet the criteria for probable Alzheimer's disease, but exhibit variations in the disease course or have a systemic illness that is sufficient to cause dementia, but is not considered to be the cause of the dementia).

The numerous varieties of dementia, and variations in patient presentations, often make diagnosis problematic. Thus, the NINCDS/ADRDA criteria are very detailed, and necessitate the thorough examination of patients with suspected Alzheimer's disease. Currently, the only way to obtain a definite diagnosis is by post-mortem histological examination of brain tissue for the presence of senile plaques.

The histopathological lesions of Alzheimer's disease include neuritic or senile plaques, neurofibrillary degeneration, and granulovacuolar neuronal degeneration. The senile plaques usually contain a core of insoluble, amyloidic extracellular material ("β-amyloid) surrounded by a halo of neurofibrillary tangles and dystrophic neurons. The primary protein component of the amyloidic core of senile plaques is a 4.2 kd amyloid β peptide (often referred to as "Aβ"). A number of other proteins have also been identified

as components of senile plaques, including anti-chymotrypsin and apolipoprotein E. The major component of the neurofibrillary tangles is an abnormally phosphorylated microtubule-associated protein referred to as "tau." The β-amyloid protein in senile plaques is a small glycoprotein (*i.e.*, a 39-43 amino acid protein, derived from the larger amyloid precursor protein), that has been detected in non-neural tissues (*e.g.*, skin, subcutaneous tissue, and intestines) and blood vessels of Alzheimer's disease patients (*See*, Cutler, *supra*). Deposits of amyloid may be detected by their ability to bind specific dyes, such as Congo red or thioflavin S. Thus, it has been further hypothesized that detection of this protein may serve as a potential diagnostic aid in the assessment of patients suffering from dementia.

Alzheimer's disease usually arises spontaneously, although genetics play a role in development of the disease. For example, the histopathologic lesions of Alzheimer's disease have been regularly observed in the brains of older patients dying of Down's syndrome (*See*, Cutler, *supra*). It is known that the β-amyloid gene resides on chromosome 21. As Down's patients have an extra copy of this chromosome, there is an increased expression of this chromosome in these patients. It has been hypothesized that increased expression of the β-amyloid gene may regulate the formation of amyloid plaques ("senile plaques") in these patients. In addition, four loci were recently identified as playing a role in the genetic susceptibility of Alzheimer's disease (*See*, Pericak-Vance and Haines (1995) *Trends Genet.* **11**:504).

Other potential risk factors for the development of Alzheimer's disease include environmental factors (*e.g.*, head trauma, smoking, and exposure to heavy metals), sociological factors (*e.g.*, depression and educational level), biological factors (*e.g.*, increasing age and hyperthyroidism), and a family history of Alzheimer's disease, Down's syndrome, or Parkinson's disease (*See*, Pericak-Vance and Haines, *supra*). Nonetheless, despite recent advances, the exact etiology and pathogenesis of Alzheimer's disease remain largely unknown.

In addition to the problems associated with diagnosing Alzheimer's disease, improvements are needed in the area of treatment. Many agents have been tested for their ability to treat the cognitive decline associated with Alzheimer's disease. For example,

various cholinergic enhancers (*e.g.*, choline and lecithin) have been tested. Unfortunately, cholinergic precursors have been shown to be not useful, although some drugs that stimulate cholinergic transmission may be helpful in some patients. One example is physostigmine, a compound that prevents the synaptic breakdown of acetylcholine. However, the overall clinical effect of this drug has not been as dramatic as initially hoped (*See*, M.A. Jenike, "Psychiatry," *in Medicine*, Scientific American, New York, (1991), pages 13 (V):1-5).

Other drugs, such as tetrahydroaminoacridine (THA or tacrine), a centrally acting anti-cholinesterase, have been tested. In a large multi-center trial of THA, liver enzyme abnormalities were reported and the preliminary results indicated that, at least at low dosages, THA is not an effective treatment of Alzheimer's disease (Jenike, *supra*). Ergoloid mesylates (Hydergine) is an extremely safe compound, and remains the most commonly prescribed drug for patients with Alzheimer's disease. However, the overall effects of the drug are at best minimal.

The mechanisms of toxicity due to accumulation of β-peptide are currently the subject of much investigation, and no definitive causes of toxicity have yet been established. (See, Benzi and Moretti (1995) *Neurobiology of Aging*, **16**:661-674). Because of this uncertainty, efforts to develop mechanism-based treatment regimens have not been possible. Compounds presently in use to treat Alzheimer's disease only serve to alleviate the systemic effects associated with the disease.

In order to study disease mechanisms and genetic-based phenomena, animals in which a foreign gene has been inserted have been described by various researchers. International Patent Application WO 96/03034 describes insertion of retroviral vectors into fish, in order to produce fish with desirable traits or to study development. Various transgenic animal models for Alzheimer's disease are described in International Patent Applications WO 93/14200, WO 93/02189, WO 94/12627, WO 94/23049, and European Patent Publication EP 653154. Typically these transgenic animals are mice or other mammals; however, β-amyloid peptide has been expressed in the nematode *Caenorhabditis elegans* (Link (1995) *Proc. Natl. Acad. Sci.* **92**: 9368). The use of *C. elegans* for investigation of mutant forms of the *C. elegans* genes mec-4 and deg-1, which

cause neurodegeneration, is described in U.S. Patent No. 5,196,333. Drawbacks to mammalian animals as model systems are the relatively long generation time, which makes mammals less desirable for high-throughput screening of potential pharmaceuticals, and the difficulty in studying the molecular processes of interest without sacrificing the animal for cell and tissue analysis.

Reporter genes are genes that encode proteins or other compounds that can be detected by a variety of methods, and which "report" the occurrence of successful introduction and expression of gene sequences. β-galactosidase and luciferase are examples of such reporter genes. Recently, the gene for green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* has been described in U.S. Patent No. 5,491,084. This gene can provide a method for indicating expression of a gene via fluorescent detection of GFP. The drawback to using GFP in mammalian systems is as described above, that is, the need to sacrifice the mammal in order to analyze the cells and tissues of interest. This drawback is especially severe in the context of large-scale screening of potential therapeutic compounds, and U.S. Patent No. 5,491,084 does not describe a system suitable for high-throughput screening of pharmaceuticals for activity against toxic proteins expressed in cells.

Thus, methods and animal systems are needed to screen drugs quickly and inexpensively for their effects on proteins and other substances associated with Alzheimer's disease. Convenient methods and animal systems for screening drugs for other neurodegenerative diseases, such as prion diseases, Huntington's disease, and amyotrophic lateral sclerosis, are also desirable.

SUMMARY OF THE INVENTION

The present invention provides methods and compositions useful for the production and use of dually transgenic animals, in particular *Caenorhabditis elegans*. In addition, in one embodiment, the present invention provides transparent animals that express a reporter gene inducible by a toxic transgene. In a preferred embodiment, the toxic transgene encodes β-peptide. Although it is not intended that the present invention be limited to any particular reporter, in a preferred embodiment, the reporter gene is green fluorescence

protein (GFP). It is contemplated that the effects or product of the reporter gene be observable (*i.e.*, the reporter gene is expressed). For example, in embodiments in which GFP is the reporter, its presence is detected using fluorescence microscopy. In addition, in these embodiments, the animals do not need to be sacrificed in order to observe the expression of the reporter gene (*i.e.*, the detection may be accomplished on living animals). If other reporters are used, other detection methods may be necessary. For example, *lacZ* expression may be detected by exposing the tissues of the animal to the substrate for the gene (*i.e.*, β-galactoside), and observing for the presence of blue dye in the tissues. However, this method requires that the animal be sacrificed in order to observe the expression of the reporter gene.

In one alternate embodiment, the genome of the transparent animal further contains a heterologous marker gene. In a preferred embodiment, the marker gene is *rol-6(su-1006)*. In the embodiments in which *rol-6(su-1006)*, expression of the marker gene is accomplished by observing the living animals for the rolling behavior that results due to the expression of the abnormal collagen gene. In contrast, wild-type animals exhibit normal locomotion and do not exhibit this rolling behavior, as their collagen is normal.

In an alternative embodiment, the first transgene of the dually transgenic non-human animal comprises pCL25. In another alternative embodiment, the second transgene of the dually transgenic non-human animal comprises pCL12. In one alternate embodiment, the genome of the transparent animal further contains a heterologous marker gene. In a preferred embodiment, the marker gene is *rol-6(su-1006)*. In the embodiments in which *rol-6(su-1006)*, expression of the marker gene is accomplished by observing the living animals for the rolling behavior that results due to the expression of the abnormal collagen gene. In contrast, wild-type animals exhibit normal locomotion and do not exhibit this rolling behavior, as their collagen is normal.

In a preferred embodiment, the transparent animal is selected from the class Nematoda. Although it is not intended that the transparent animal of the present invention be limited to any specific animal, in a particularly preferred embodiment, the transparent animal is *Caenorhabditis elegans*.

In an alternate embodiment the cells of the transparent animal display toxicity resulting from the accumulation of β-peptide within the cells. In one preferred embodiment, the genome of the transparent animal comprises SEQ ID NO:5 and SEQ ID NO:8.

The present invention also provides methods for producing dually transgenic non-human animal comprising: providing: a first and second non-human animal: a first transgene comprising β-peptide: and a second transgene, comprising a reporter: introducing the first transgene into the genome of the first non-human animal to produce a first transgenic animal, and introducing the second transgene into the genome of the second non-human animal to produce a second transgenic animal: and mating the first transgenic animal with the second transgenic animal to produce a dually transgenic animal, wherein the β-peptide and the reporter are expressed.

In one embodiment of the methods of the present invention the dually transgenic non-human animal is transparent. In a preferred embodiment, the transparent animal is a nematode, while in a particularly preferred embodiment, the animal is *Caenorhabditis elegans*.

In an alternative embodiment, the first transgene of the dually transgenic non-human animal comprises pCL25. In another alternative embodiment, the second transgene of the dually transgenic non-human animal comprises pCL12. In one alternate embodiment, the genome of the transparent animal further contains a heterologous marker gene. In a preferred embodiment, the marker gene is *rol-6(su-1006)*. In the embodiments in which *rol-6(su-1006)*, expression of the marker gene is accomplished by observing the living animals for the rolling behavior that results due to the expression of the abnormal collagen gene. In contrast, wild-type animals exhibit normal locomotion and do not exhibit this rolling behavior, as their collagen is normal.

The present invention also comprises methods for testing compounds for anti-toxic effects, comprising: providing a dually transgenic non-human animal expressing a toxic transgene and a reporter: a composition comprising a test compound in a form suitable for administration such that the compound is bioavailable in the cells of the animal; and administering the test compound to the non-human animal. The toxic transgene induces

expression of the reporter gene. In a preferred embodiment, the toxic transgene is β-peptide and the reporter gene is GFP. In one embodiment, the method further comprises the step of measuring a reduction or increase in the expression of the reporter by the dually transgenic non-human animal and thereby identifying a compound as therapeutic. In a particularly preferred embodiment of the methods, the compounds inactivate the β-peptide expressed by the dually transgenic animal.

In one embodiment of the methods for testing compounds for β-peptide toxicity, the dually transgenic non-human animal is transparent. In a preferred embodiment, the transparent animal is a nematode, while in a particularly preferred embodiment, the animal is *Caenorhabditis elegans*. In an alternative embodiment, the first transgene of the dually transgenic non-human animal comprises pCL25. In another alternative embodiment, the second transgene of the dually transgenic non-human animal comprises pCL12. In one alternate embodiment, the genome of the transparent animal further contains a heterologous marker gene. In a preferred embodiment, the marker gene is *rol-6(su-1006)*, which produces effects described herein.

In one alternative embodiment, one or more test compounds are tested for their ability to counter the toxic effects of transgene product. It is also contemplated that the test compounds will be tested for their ability to prevent the expression of the toxic transgene, for example, β-peptide.

In alternative embodiments, dually transgenic animals in which the expression of, or effects of transgene toxicity, are reduced or eliminated by the test compounds, are mated. The progeny of these matings are also then tested for the effects of test compounds on the expression of, or effects of toxicity. In yet another embodiment, the progeny of these matings are used in other assay systems for the identification of therapeutic compounds.

In yet another alternative embodiment, the dually transgenic animals of the present invention are used to identify methods suitable for the diagnostic testing of Alzheimer's disease or pathology due to other diseases. Thus, the dually transgenic animals of the present invention are used to develop assays suitable for use in humans or animal models of Alzheimer's disease.

DESCRIPTION OF THE FIGURES

Figure 1 is a diagram illustrating the construction of dual transgenic animals expressing β-amyloid peptide and an hsp/GFP reporter.

Figure 2 shows the DNA sequence (SEQ ID NO:5) and restriction map for pCL12.

Figure 3 is a graphic map of pCL12.

Figure 4 shows the DNA (SEQ ID NO:6), and amino acid sequence (SEQ ID NO:7) of pCL12 from nucleotide 1071 through 1253 (*i.e.*, the β-(1-42) nucleic acid and amino acid sequence).

Figure 5 shows the DNA sequence (SEQ ID NO:8) and restriction map for pGFP-TT.

Figure 6 is a graphic map of pGFP-TT.

Figure 7 shows the DNA sequence (SEQ ID NO:9) and restriction map for pCL25.

Figure 8 is a graphic map of pCL25.

Figure 9 shows the DNA sequence (SEQ ID NO:10) of *rol*-6.

Figure 10 shows the amino acid sequence (SEQ ID NO:11) of *rol*-6.

DESCRIPTION OF THE INVENTION

The present invention provides methods and compositions useful for the production and use of transgenic animals. The methods and transgenic animals of the invention also provide an efficient and effective system for screening drug effective in ameloriating the effects of toxic gene products.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, cell biology and recombinant DNA, which are within the skill of the art. *See, e.g.*, Sambrook, Fritsch, and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, (F.M. Ausubel et al. eds., 1987); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (M.J. McPherson, B.D. Hames and G.R. Taylor eds., 1995); ANIMAL CELL CULTURE (R.I.

Freshney. Ed., 1987); and ANTIBODIES: A LABORATORY MANUAL (Harlow et al. eds., 1987).

Animals

The methods and systems described herein can be practiced with any non-human animal. In a preferred embodiment, the animal used is transparent for at least part of its life cycle, for example, oocytes (*Xenopus* or others); larvae; pupae, fish (zebrafish or others); and nematodes, such as *C. elegans*. *C. elegans* is particularly preferred because it is transparent throughout its life and has been extensively studied as a model to determine developmentally regulated gene expression, as well as pattern formation. The visibility of all of the animals' cells throughout their life cycles, has led to the complete determination of the cell lineages, and a detailed description of the morphogenic changes that occur during *C. elegans* development. Rapid methods for introduction of cloned DNA into the *C. elegans* germ line has provided means to study molecular function and expression *in vivo*, as the effects of the gene expression are usually readily visible in the transparent worms.

Reporter Molecules

The methods and compositions of the present invention make use of reporter genes to monitor gene activity. Visible and quantifiable reporter genes are known and described in the art. Successfully used reporter molecules in gene fusion vectors in studies with *C. elegans* include the *E. coli lac*Z coding region (*See e.g.*, A. Fire *et al.*, (1990) *Gene* **93**:189-198), and the *Aequora victoria gfp* coding region (*See e.g.*, M. Chalfie *et al.*, (1994) *Science* **263**:802-805; D.C. Prasher *et al.*, (1992) *Gene* **111**:229-233; and Genbank Accession #M62654), which produces an intrinsically fluorescent protein. Although various fusion expression vectors have been used and reported in the literature, problems have been encountered (*See e.g.*, A. Fire, "Fire Lab Vector Kit"--June 1995). For example, with *lac*Z, there have been many reports in which it was not possible to correlate transgene expression patterns with physiological expectations, or only weak correlations have been possible. Ectopic expression is often frequently seen with short promoters, and occurs

most prominently in the gut and pharynx of the affected animals. This may be due to weak promoter and/or enhancer signals in the vectors. Expression pattern deficits of *lacZ* fusions have been classified into three groups. In the first, transgenes are expressed in the correct tissue, but mosaic expression is observed (*i.e.*, only a subset of the cells stain during the detection methods). This has been observed even with integrated high-copy transgenes. In the second group, expression in a single tissue or cell population is not seen with the transgene. Finally, there has been the failure of transgene fusion constructs to show expression in the pre-12 cell embryo, or in any embryonic, larval, or adult germ line.

The *gfp* fusion vectors were developed as an alternative to the *lacZ* markers, but have been available for a much shorter time (*See e.g.*, Chalfie *et al.*, (1994), *supra*; U.S. Patent No. 5,491,084). Initial reports indicate that the fluorescence pattern appears to be more restricted than that exhibited by equivalent *lacZ* fusion. In addition, there are still problems associated with the germline expression of the transgenes and ectopic expression. Indeed, in some cases, the problems with ectopic expression have been exacerbated by these "improvements."

The present invention overcomes many of these problems described in the art. The tranparent animals and dually transgenic animals described herein provide easily visible, stably expressed systems in which the toxic transgene and reporter gene are expressed similarly to endogenous chromosomal gene expression. These animals provide an excellent system for screening compounds having effects on the toxic transgene. The present invention, therefore, provides model systems for the study of human diseases and methods of identifying therapeutic compounds using these animal systems.

Inducible Promoters

As discussed above, the reporter genes are operably linked to an inducible promoter. The promoter is induced by the toxic transgene, for example by the gene product of the toxic transgene. It is also contemplated that the toxic transgene can act to induce the promoter indirectly, for example by disrupting other cellular proteins or functions. Suitable inducible promoters are available and can be readily determined by those skilled in the art. Non-limiting examples of promoters which are induced by

"stresses" include the metallothionein gene promoter (*mtl*-1 or *mtl*-2; *e.g.*, Genbank Accession #M92910, #M11794, #X00504, and #X00953), and the *C. elegans* amyloid precursor protein (APP) homolog *apl*-1 gene promoter (*See e.g.*, Daigle and Li, Proc. Natl. Acad., Sci., 90:12045-12049 [1993]). Other potential promoters include those from other inducible heat-shock genes; at least one of the known *C. elegans hsp*70 genes is strongly heat-inducible. It is also contemplated that promoters from genes known to be up-regulated under stress conditions in other systems (*e.g.*, superoxide dismutase, catalase, glutathione reductase, etc.) may also be useful.

In one embodiment, as described in the Examples below, a strong muscle-specific promoter was used to express a potentially secretable form of the β-peptide, so as to generate significant extracellular levels of β-peptide. This was accomplished in order to mimic the situation that may exist in the human brain, and allow observations of cell-external neurotoxicity. The present inventor has also demonstrated that the promoter must be chosen so that enough β-peptide is produced to cause physiological effects, but not so much to kill the animals.

The *unc*-54/β-(1-42) minigene was constructed with a modified signal sequence that has been previously shown to allow secretion of a *her*-1 protein product that is ectopically expressed in muscle cells (M.D. Perry *et al.*, (1993) *Genes Dev.*, 7:216-228). The development of *C. elegans* transgenic for expression of β-amyloid was previously described by the inventor (Link, (1995) *Proc. Natl. Acad. Sci.*, 92:9368-9572).

However, in early experiments, it was observed that β-peptide deposits were not convincingly detected outside of the muscle cells, when tested with the antibodies described in Example 4, below. While an understanding of the mechanism is not necessary for the practice of the present invention, it is apparent that the majority of β-peptide expressed by the transgenic worms is retained in the muscle cells and is responsible for the pathology observed in the muscle cells.

It is also contemplated that β-peptide expression will be directed to other tissues, through utilization of appropriate promoters. For example, it is contemplated that animals expressing β-peptide in the intestine may be particularly useful to analyze compounds such

as drugs for their effect on β-peptide, as these cells readily take up exogenous compounds administered orally.

Transgenes

The transgene element of the present invention can be any sequence which is able to induce the promoter operably linked to the reporter gene. As noted, the transgene is "toxic" in the sense that it disrupts cellular function in some way. Preferably, the transgene encodes a protein that is toxic to the host cells and/or organism in that it causes deleterious effects to the host, for instance interfering with the hosts ability to survive and/or grow. The toxic transgenes encode proteins that disrupt cellular function directly (*e.g.*, the gene product is toxic) or indirectly (*e.g.*, the sequence of the transgene disrupts cellular function by some mechanism other than its gene product). Toxic proteins are distinguishable from simple chemical toxins (*e.g.*, heavy metals and the like) by their antigenicity and higher molecular weight.

In one embodiment, the toxic transgene comprises a gene encoding for an amyloidic protein, for example beta amyloid peptide, prion protein variants, transthyretin variants, gelsolin variants, cystatin variants, lysozyme variants and the like. In another embodiment, the transgene encodes a protein containing polyglutamine resulting from triplet-repeat expansion such as huntingtin (a protein that has been implicated in Huntington's Chorea), ataxin-1 or ataxin-2. Alternatively, proteins associated with inherent amyelotopic lateral sclerosis (ALS) for example, superoxide dismutase 1 variants and over-expressed neurofilament protein, can be used. It will be understood that the transgene can encode for an entire toxic protein or, alternatively, a functional (*i.e.* toxic) fragment.

Transgenic Animals

In one aspect, the present invention includes a dually transgenic non-human animal. In one embodiment, the development of these dually transgenic animals involves the production of two *C. elegans* lines with a single transgene in each line, designated as CL2005 and CL2070. *C. elegans* was chosen due to the observation that the toxic effect of

β-amyloid has been reported to occur in these animals, in addition to humans. Line CL2005 exhibited muscle-specific expression of human β peptide, while line CL2070 exhibited stress-inducible expression of GFP. GFP was incorporated into the present invention as it provides a marker (*i.e.*, "reporter") that is readily visible in living worms. Thus, dually transgenic animals both express β-peptide in their muscle cells and under appropriate conditions for observation (*i.e.*, fluorescence microscopy), exhibit green fluorescence in these cells.

Mating of these lines resulted in the production of dually transgenic animals, in which the presence of β-amyloid was easy to detect, due to the fact that the dually transgenic animals expressing β-amyloid glowed green. Thus, the present invention provides an easily detectable method for the specific expression of β-peptide in dually transgenic animals. This ease of detection provides great advantages for the development of methods to analyze the effects of β-amyloid *in vivo*. In addition, because the dually transgenic animals are not killed in order to detect the presence of β-amyloid, as would be required if other reporters (*e.g.*, *lacZ* or luciferase) were used, the animals may be useful for screening compounds for their effects on β-amyloid in an *in vivo* situation. Thus, the use of an reporter which is detectable *in vivo* provides significant advantages over currently available methods.

Other Advantages of Transgenic *C. elegans*

The present invention also includes methods of screening compounds for their ability to prevent or inhibit toxicity due to the toxic transgene. In these drug screening embodiments, the methods and animals described above can be used to identify potential therapies. For example, in assays to develop compounds to block the expression of β-amyloid or counter its toxic effects, transgenic worms (*e.g.*, green fluorescing worms with muscle-specific β-peptide expression), would be observed for their loss of fluorescence. In assays where the compound is effective in countering the expression of β-peptide or effects of β-peptide, the animals would no longer glow green. It is easy to visually screen and quantify (using commercially available equipment) for animals that have lost the ability to glow, and then use them to detail the interaction between the compound and the toxic

transgene. The present invention provides a fast and efficient screening system, for example by using commercially available equipment for assaying multiple compounds at once.

Thus, the transgenic animals of the present invention expressing a detectable reporter gene triggered by a toxic transgene provide animal models for human diseases. For instance, a reporter gene (such as GFP) induced by expression of β-peptide provides an animal model for human Alzheimer's disease. It is also not intended that the present invention be limited to animals expressing β-peptide in conjunction with GFP. It is contemplated that expression of other proteins associated with dementia and/or Alzheimer's or other diseases would also find use in conjunction with GFP in *C. elegans* as well. For example, it is contemplated that proteins (including mutated versions of proteins) such as the presenilins associated with Alzheimer's disease (*See e.g.*, J. Marx (1996) *Science* **274**:1838-1840) will be used in conjunction with GFP and full-length APP in dually transgenic *C. elegans*. Other toxic proteins which provide animal models for other human diseases are described herein.

These transgenic animals also provide a means to screen compounds for their ability to decrease or eliminate the toxicity. Examples of compounds suitable for testing using the transgenic animals of the present invention include, but are not limited to, Congo Red, tumor necrosis factor (TNF), estrogen, tacrine (9-amino-1,2,3,4-tetrahydroacridine), dihydroepiandrosterone (DHEA), compounds that inhibit ApoE4, and others, commercially available from suppliers such as Sigma. Compounds to be tested for anti-toxic activity are administered to the same number of dually transgenic animals (*e.g.*, generated using the methods described in the Examples, below) from the control group and the treatment group, and the presence or absence of reporter gene used as a measure of efficacy.

The compounds being tested can be administered using any suitable route (*e.g.*, oral, parenteral, controlled-release transdermal methods, and implants, etc.). In one preferred route, the compounds to be tested are suspended in the growth media provided to the worms. Generally speaking, the route of administration will depend on the stability of the compound, the susceptibility of the compound to "first pass" metabolism, the

concentration needed to achieve a therapeutic effect, and the like. Following initial screening, a compound that appears promising (*i.e.*, which increases the number of worms which display reduced β-peptide toxicity) is further evaluated by administering various concentrations of the compound to additional transgenic animals in order to determine an approximate therapeutic dosing range.

Another screening method involves the crossing of the transgenic worms of the present invention with other transgenic worms. The animals are observed after treatment, in the presence and absence of the test compound(s), with the effects on the toxic transgene being gauged either by crude survival or the presence/absence of the reporter gene. It is also contemplated that the methods of the present invention be modified so as to provide means to analyze disease-related proteins believed to have dominant toxic effects, by substituting appropriate sequences for the β-peptide sequences used as described in the Examples. For example, proteins such as transthyretin (known to be associated with familial amyloid polyneuropathy; *See e.g.*, Christmason *et al.*, (1991) *FEBS* **281**:177-180; Genbank Accession #D00096), and variant superoxide dismutase (known to be associated with familial amyotrophic lateral sclerosis [Lou Gehrig's disease]), prion proteins, A4 amyloid protein (*See e.g.*, Ponte *et al.*, (1988) *Nature* **331**:525-527; Salbaum *et al.*, U.S. Patent No. 5,151,508, herein incorporated by reference), APP (*See e.g.*, Kitaguchi *et al.*, (1988) *Nature* **331**:530-532; Sata *et al.*, EP Appln. 94117512.7; Scott *et al.*, WO 9412627; Wadsworth *et al.*, WO 9314200; Gearhart *et al.*, WO 9423049; and Neve *et al.*, WO 9302189, all of which are herein incorporated by reference), other amyloidic proteins (*e.g.*, variant lysozymes and amylin peptide), and other proteins associated with neuronal degeneration (*See e.g.*, U.S. Patent No. 5,196,333), will be used in the methods of the present invention by substituting the appropriate nucleic acid sequences encoding the protein of interest for the β-peptide described in Example 1. It is also contemplated that animals transgenic for dual proteins (*e.g.*, β-peptide in combination with another protein, such as transthyretin) will be used. Thus, the present invention provides the means to analyze the effects of numerous genes and proteins *in vivo*.

Furthermore, the transgenic animals of the present invention provide distinct advantages over other transgenic animals currently used to analyze diseases such as

Alzheimer's disease. Due to the short gestation period of *C. elegans*, transgenic animals can be produced much more rapidly than when mammals, such as mice are used. For example, transgenic mice overexpressing a 695 amino acid isoform of β-amyloid precursor did not show learning and memory impairment until they were 9-10 months of age (Hsiao *et al.*, Science 274:99-102 [1996]). In contrast, the transgenic animals of the present invention express β-peptide and may be manipulated at a very early age, even in embryonic stages.

In addition, because the transgenic animals are easy to select based on the presence of the green fluorescence, invasive procedures such as surgery, necessary to analyze the effects of the toxic transgene in other animals (*e.g.*, mice) are avoided.

Finally, the apparent toxicity of β-peptide in transgenic animals was found to be temperature-dependent. Transgenic animals maintained at 25.5°C were significantly more sick than those maintained at 16°C. The animals maintained at the higher temperature became paralyzed more quickly, failed to eat or grow, failed to show normal egg-laying, and many died before reaching adulthood. However, wild-type animals grow well at either of these temperatures. The CL2005 parental line was temperature-sensitive for viability, as it was not possible to propagate this line at the elevated temperature. In the CL2070 parental line, 25.5°C maintenance was not sufficient to induce the hsp/GFP construct. The hsp/GFP response was similarly temperature-dependent. It was difficult to detect GFP when the dual transgenic animals were raised at 16°C, but GFP induction was dramatic when the animals were raised at 25.5°C.

These temperature dependency observations were exploited by the propagation of animals at 16°C, and then upshifting them to 25.5°C, when GFP induction was desired. This effect also has potential benefits in the analysis of compounds, as the animals can be pre-incubated in the presence of drugs or other compounds at the lower temperature before the upshift to the higher temperature, in order to ensure that the presumed protective effect of the drug was in place, prior to the strong induction of β-peptide toxicity. Importantly, these assays may be completed within one day and are suitable for rapid methods (*e.g.*, the use of a microtiter format and a plate fluorimeter), so that literally thousands of compounds may be tested simultaneously.

In sum, the present invention provides methods and compositions useful as animal models for disease, as well as providing methods and compositions for disease therapy and prevention. The animal testing may be supplemented and confirmed by testing on human subjects. However, the transgenic animals of the present invention allow the testing of a large number of compounds, both various methods, including those known in the art.

DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

The term "non-human animal" includes vertebrates such as rodents, arthropods, insects (*e.g.*, Diptera), fish (*e.g.*, zebrafish), non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. The non-human animal can be used at any stage in its development, for example oocyte, fetal, larval, pupal stages or the like. Preferred non-human animals are those that are transparent, such as certain nematodes, oocytes, larvae and fish.

As used herein, the term "transparent" is used in references to animals through which light will be transmitted. However, it is not intended that the amount of light transmittance be limited to any particular amount. For example, an animal is transparent so long as at least some light may be transmited through its body. Non-limiting examples of transparent animals are larval stages of some animals (*e.g.*, flies) and oocytes (*e.g.*, *Xenopus* oocytes). Particularly preferred non-human animals are selected from the nematodes (*i.e.*, any animal in the Class Nematoda), most preferably *Caenorhabditis elegans*. However, it is contemplated that other transparent animals, such as zebrafish will be useful in the present invention. For example, transgenic zebrafish have been produced (*See e.g.*, Lin *et al.*, PCT Publ. WO9603034; incorporated herein by reference).

The "non-human animals having a genetically engineered genotype" of the invention are preferably produced by experimental manipulation of the genome of the germline of the non-human animal. These genetically engineered non-human animals may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into an embryonal target cell or integration into a chromosome of the somatic and/or germ line cells of a non-human animal by way of human

intervention, such as by the methods described herein. The process by which a DNA molecule becomes stably incorporated into another genome is referred to as "stable integration." Non-human animals which contain a transgene are referred to as "transgenic non-human animals". A transgenic animal is an animal whose genome has been altered by the introduction of one or more transgenes.

The term "transgene" as used herein refers to a foreign gene that is placed into an organism by introducing the foreign gene into gonadal cells, embryonic cells, newly fertilized eggs, or early embryos. The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) which is introduced into the genome of an animal by experimental manipulations and may include gene sequences found in that animal so long as the introduced gene contains some modification relative to the naturally-occurring gene. A preferred foreign gene is the β-amyloid gene (*e.g.*, β-peptide), or fragments thereof. A "toxic transgene" refers to a foreign gene which disrupts cellular function in some way. For example, the toxic transgene may produce a gene product (*i.e.*, protein) which is toxic to the cell or organism. Alternatively, the transgene may disrupt other cellular proteins, or act as a DNA binding site without producing a protein.

As used herein, the term "toxicity" refers to the production of toxic effects by any compound or substance. For example, while it is not necessary to the understanding and use of the present invention, β-peptide toxicity may arise due to the accumulation of β-peptide molecules. There are numerous manifestations of toxicity that may occur. For example, the toxicity associated with β-peptide may be manifested as negative effects on muscle cells of dually transgenic animals.

The terms "promoter element," "promoter," or "promoter sequence" as used herein, refer to a DNA sequence that is located at the 5' end of (*i.e.*, precedes) a gene in a DNA polymer and provides a site for initiation of the transcription of the gene into mRNA. An "inducible promoter" is a promoter that is triggered by certain signals within the cell, for instance binding of a transcription factor, stress, heat or the like.

The term "reporter gene" as used herein refers to genes that encode proteins or other compounds that can be detected by a variety of methods. These genes "report" the occurrence of successful introduction and expression of sequences such as transgenes.

Non-limiting examples of reporter genes include antibiotic resistance genes, genes encoding enzymes and genes encoding other detectable proteins. Expression of the reporter gene is detected using methods known in the art. In a preferred embodiment, the reporter gene is a GFP gene, although *lacZ* β-galactosidase gene, or any other reporter system may be used to detect the successful production of transgenic animals. In particularly preferred embodiments, the reporter is a compound or protein which may be present or expressed within living animals. That is, it is not necessary to sacrifice the animal in order to detect the presence of the reporter. The type of the reporter gene system used is not critical to the invention, and it is contemplated that any system suitable for use with the transgenic animals of the present invention will be used.

As used herein, the terms "dual trangenic" and "dually transgenic" refer to animals or cells in which more than one transgene have been introduced. For example, the term is used in reference to cells which contain the sequences encoding β-peptide, and sequences encoding a reporter (*e.g.*, GFP). However, it is not intended that the number of transgenes in the dually transgenic animals of the present invention be limited to two. For example, the transgenic animals may also contain another one or more marker genes (*e.g.*, *rol-6* sequences), in addition to β-peptide and reporter sequences. In the embodiments in which *rol-6(su-1006)*, expression of the marker gene is accomplished by observing the living animals for the rolling behavior that results due to the expression of the abnormal collagen gene. In contrast, wild-type animals exhibit normal locomotion and do not exhibit this rolling behavior, as their collagen is normal.

The transgenic animals of the present invention are preferentially generated by introduction of the targeting vectors into gonad cells. Transgenes can be efficiently introduced into the cells by DNA transfection using a variety of methods known to the art, including electroporation, calcium phosphate co-precipitation, protoplast or spheroplast fusion, lipofection, and DEAE-dextran-mediated transfection. Transgenes may also be introduced into cells by retrovirus-mediated transduction or by micro-injection. In one preferred embodiment, the transgenes are injected into gonads of *C. elegans* as described by Mello *et al.* (1991) *EMBO J.*, **10**:3959-3970. Alternative methods for the generation of transgenic animals containing an altered gene are known to the art. For example,

embryonal cells at various developmental stages can be used to introduce transgenes for the production of transgenic animals. Different methods are used depending on the stage of development of the embryonal cell.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, and end-labeling or PCR amplification using a labeled nucleotide. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. The term may also be used in reference to proteins. For example, a variety of protocols which employ polyclonal or monoclonal antibodies specific for the β-peptide protein product are known in the art (*See*, the Examples). These antibodies can be used as markers for the expression of proteins. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS).

The term "gene" refers to a DNA sequence that comprises coding sequences necessary for the production of a polypeptide or precursor. In some instances, a gene can also include control sequences. The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired enzymatic activity is retained.

The term "gene of interest" refers to any gene, the manipulation of which may be deemed desirable for any reason, by one of ordinary skill in the art.

The term "wild-type" refers to a gene or gene product which has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product which displays modifications in sequence and or functional properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated: these are

tified by the fact that they have altered characteristics when compared to the wild-type or gene product.

The terms "targeting vector" or "targeting construct" refer to oligonucleotide nces comprising a gene of interest flanked on either side by regulatory sequences. ably, the targeting vector is capable of homologous recombination such that the gene rest is integrated by recombination.

As used herein, the terms "vector" and "vehicle" are used interchangeably in ce to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The terms "expression vector" or "expression cassette" as used herein, refers to a inant DNA molecule containing a desired coding sequence and appropriate nucleic uences necessary for the expression of the operably linked coding sequence in a r host organism. Nucleic acid sequences necessary for expression in prokaryotes nclude a promoter, an operator (optional), and a ribosome binding site, often along r sequences. Eukaryotic cells are known to utilize promoters, enhancers, and n and polyadenylation signals.

e terms "in operable combination," "in operable order," and "operably linked" as n, refer to the linkage of nucleic acid sequences in such a manner that a nucleic ule capable of directing the transcription of a given gene and/or the synthesis of rotein molecule is produced. The term also refers to the linkage of amino acid n such a manner so that a functional protein is produced.

term "tissue-type specific" as it applies to a promoter, refers to a promoter that f selectively directing expression of a gene in a specific tissue. Similarly, the e-specific promoter in the method of the present invention does not require cificity. In general, the requisite specificity is found where a plurality (or, bly, a majority) of cells in one tissue type express a gene of interest, while .g., greater than 80%, and preferably greater than 90%, and more preferably 5%), of the cells in other tissue types do not. In one embodiment of the ion, the strong muscle promoter *unc*54 was used (*See*, Example 1).

d herein the term "portion" when in reference to a gene refers to fragments of fragments may range in size from a few nucleotides to the entire gene

sequence minus one nucleotide. Thus, "an oligonucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" is therefore a substantially purified polynucleotide.

As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis disclosed in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization

with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; and/or incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (*e.g.*, genes) for the construction of targeting vectors, transgenes, etc.

As used herein, the terms "PCR product" and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

The term "Central Nervous System" refers to the "spinal cord" and the "brain." The spinal cord comprises white areas and grey areas. The grey area contains nerve cell bodies, whereas the white area is essentially comprised of myelinated nerves. The brain, which is also known as the encephalon, is that portion of the cerebrospinal axis which is contained in the cavity of the cranium. The brain comprises the two cerebral hemispheres, the inter-brain, the mid-brain, the pons Varolli and cerebellum, and the medulla oblongata. The two hemispheres together with the parts derived from the thalamencephalon form the forebrain. The two cerebral hemispheres are separated by the longitudinal fissure and also comprise the bylvian fissure, the fissure of Rolando, and the parieto-occipital fissure. The lobes on the external surface of the brain comprise the frontal lobe, the parietal lobe, the occipital lobe, and the temporal lobe. Placed along the middle line of the brain are, among others, the rostrum and peduncles of corpus callosum, lamina cinera, optic commissure and the pituitary body. On each side of the middle line lies the frontal lobe, olfactory lobe and the hemisphere of cerebellum.

The terms "neuron," "neural cell," and "nerve cell" are used interchangeably to refer to a cell which is located in the nervous system. Nerve cells are composed of the nerve cell body (perikaryon), one or more dendrites, and an axon. Neurons can be classified according to the number of processes originating from the cell body. Thus,

unipolar neurons have a single process, bipolar neurons have one axon and one dendrite, while multipolar neurons (which are the most common) comprise more than two processes. The term "neuron" comprises cholinergic neurons and sensory neurons. As used herein, the term "cholinergic neuron" means a neuron in the Central Nervous System (CNS) and in the Peripheral Nervous System (PNS) whose neurotransmitter is acetylcholine. As used herein, the term "sensory neuron" includes a neuron which is responsive to environmental cues (*e.g.*, temperature and movement) from, for example, the skin, muscle and joints of a mammal.

The term "nerve" refers to two or more neurons arranged in linear sequence such that the axon of one neuron establishes a structural and functional link with the dendrite of a second neuron to form a "synapse."

The term "compound" refers to any chemical entity, pharmaceutical, drug, and the like that can be used to treat or prevent a disease, illness, sickness, or disorder of bodily function. Compounds comprise both known and potential therapeutic compounds. A compound can be determined to be therapeutic by testing using the testing methods of the present invention (*i.e.*, a "test compound"). A "known therapeutic compound" refers to a therapeutic compound that has been shown (*e.g.*, through animal trials or prior experience with administration to humans) to be effective in such treatment or prevention. In other words, a known therapeutic compound is not limited to a compound efficacious in the treatment or prevention of neural related disorders.

A compound is said to be "in a form suitable for administration such that the compound is bioavailable in the blood of the animal" when the compound may be administered to an animal by any desired route (*e.g.*, oral, intravenous, subcutaneous, intramuscular, etc.) and the compound or its active metabolites appears in the blood of the animal in an active form. Administration of a compound to a pregnant animal may result in delivery of bioavailable compound to the embryonic progeny of the animal.

The "wild-type β-amyloid" or "β-peptide" gene and gene product refers to the nucleotide and amino acid sequences provided in SEQ ID NOS:6 and 7, respectively. Those skilled in the art will be well aware that certain modifications of SEQ ID NOS:6 and 7 can be made which will not interfere with the production of a polypeptide having an

activity indistinguishable from that of the wild-type β-amyloid; the present invention specifically contemplates these variant forms of β-amyloid. A "variant" of the β-peptide is defined as an amino acid sequence that differs by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (*e.g.*, replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes, *e.g.*, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNAStar software.

The term "an oligonucleotide sequence comprising at least a portion of a human β-amyloid gene" refers to a polynucleotide sequence (*i.e.*, a nucleic acid sequence) containing a nucleotide sequence derived from a human β-amyloid gene. This sequence may encode a portion or all of the β-amyloid protein; alternatively, this sequence may contain non-coding regions derived from the β-amyloid gene or a combination of coding and non-coding regions. The oligonucleotide may be RNA or DNA and may be of genomic or synthetic origin.

As used herein the term "portion" when in reference to a gene refers to fragments of that gene. The fragments may range in size from 10 nucleotides to the entire gene sequence minus one nucleotide. Thus, "an oligonucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

An animal whose genome comprises a "heterologous marker gene" is an animal whose genome contains a marker gene not naturally found in the animal's genome. In one preferred embodiment of the present invention, the heterologous marker gene is a mutant collagen gene, such as the *rol-6(su-1006)* gene. However, it is intended that other marker genes will be used with success in the present invention, including other mutant collagen genes, as well as other marker genes commonly known to those in the art.

As used herein, the term "diagnostic assay" refers to methods for the diagnosis of disease, illness, and/or pathology. It is intended that the term encompass any methods for

diagnosis, including, but not limited to assays based on immunoreactivity (*e.g.*, radioimmunoassays, fluorescence immunoassays, enzyme immunoassays), histochemistry, dye retention or binding (*e.g.*, fixing of dyes such as Congo Red), nucleic acid based diagnostic methods (*e.g.*, identification of nucleic acid sequences associated with disease or pathology), etc.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μM (micromolar); N (Normal); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); μg (micrograms); ng (nanograms); l or L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); °C (degrees Centigrade); U (units), mU (milliunits); min. (minutes); sec. (seconds); % (percent); kb (kilobase); bp (base pair); PCR (polymerase chain reaction); β-(1-42) amino acids 1-42 of β-amyloid peptide; Tris (tris(hydroxymethyl)-aminomethane); BSA (bovine serum albumin); Fisher (Fisher Scientific, Pittsburgh, PA); Sigma (Sigma Chemical Co., St. Louis, MO.); Promega (Promega, Corp., Madison, WI); Perkin-Elmer (Perkin-Elmer/Applied Biosystems, Foster City, CA); Senetek (Senetek, PLC, Maryland Heights, MO); Boehringer Mannheim (Boehringer Mannheim, Corp., Indianapolis, IN); Stratagene (Stratagene, Inc., La Jolla, CA); and NEB (New England Biolabs, Beverly, MA). Unless otherwise indicated, the restriction enzymes used in these Examples were obtained from NEB. *C. elegans* is available from the Caenorhabditis Genetics Center, at the University of Minnesota, St. Paul, MN.

EXAMPLE 1

Assembly of the β-(1-42) Minigene

1-42) minigene used in subsequent experiments (*See*,
ic *C. elegans* engineered to express amyloidic human
escribed below, these animals contained constructs in
54 promoter/enhancer of *C. elegans* drove the expression of
lerived from human β-amyloid cDNA clones, as described
ıman β-amyloid peptide in transgenic *Caenorhabditis*
, 92:9368-9372 [1995]).

ng the 42 amino acid β-amyloid peptide derived from
ın cDNA was assembled. The artificial signal peptide
52.81 was amplified under standard conditions, using
"GGCCAAAGGACCC-3')(SEQ ID NO:1), and "SP-down"
CCAGCAAGAT-3')(SEQ ID NO: 2), cleaved with *Nhe*I
ıd inserted between the unique *Nhe*I and *Kpn*I sites of
: construct "pCL2." This process resulted in a re-
e, such that the signal peptide cleavage site, as predicted by
von Heijne, Nucl. Acids Res., 14:4683-4690 [1986]),
Gly-Thr dipeptide encoded by the *Kpn*I site.

ding amino acids 1-42 of β-amyloid, and which contained
plified using standard methods, from human β-amyloid
4T4B (P. Ponte *et al.*, Nature 331:525-527 [1988]), by
GGGGTACCGATGCAGAATTCCGACATGA-3') (SEQ
5'-CCCGAGCTCACGCTATGACAACACCGCCAA-
cation product was cleaved with *Kpn*I and *Sac*I, and
ıI and *Sac* I sites of pCL2, to generate "pCL3."

42) minigene fragment was removed from this plasmid by
d inserted between the unique *Nhe*I and *Sac*I sites of
The sequence of the β-(1-42) minigene was confirmed by
coding strand only, by techniques known in the art. The
):5), is shown in Figure 2. A graphic map of pCL12 is
ıows the DNA and amino acid sequences (SEQ ID NOS:

6 and 7, respectively) of pCL12 from nucleotide 1071 through 1253 (*i.e.*, the β(1-42) nucleic acid and amino acid sequences) present in the construct.

EXAMPLE 2

Construction of the hsp/GFP Reporter

ımple, the expression vector ("pCL25") containing the hsp/GFP reporter :xpression of the GFP-TT gene was constructed. This vector was used in :ments (*See*, Example 3) to produce transgenic *C. elegans* was '-TT is a modified form of GFP that contains Ser65Thr and Ile167Thr . relative to the sequence of the wild-type GFP). The substitutions present ce a protein which results in the production of a much brighter ı the wild-type GFP.

ɔs constructed as described below. A 431 bp *Hin*dIII-*Bam*HI fragment p16-2 promoter was excised from the plasmid "pPD49.78," by digestion *Hin*dIII. This 431 bp *Hin*dIII-*Bam*HI fragment was inserted between the de #1) and *Bam*HI (nucleotide #31) sites of the GFP-TT gene in the ΓT," (available from Yishi Jin, at the University of California, Santa :igase (Promega). The pGFP-TT plasmid contains the coding regions for i into a Tu61 backbone. The DNA sequence of pGFP-TT (SEQ ID NO: 8) re 5. A schematic map of pGFP-TT is shown in Figure 6, in which unique ıre shown. The pGFP-TT plasmid was digested with *Hin*dIII and *Bam*HI, ragment containing the hsp16-2 promoter was inserted, in order to ss-inducible expression vector "pCL25." The DNA sequence of pCL25 and restriction map are shown in Figure 7. Figure 8 provides a schematic n which unique restriction sites are shown.

EXAMPLE 3

Construction of Transgenic Animals

In this Example, the transgenic parent animals were produced, with one line expressing the β(1-42) minigene (designated as "CL2005") and the other line expressing the hsp/GFP reporter (designated as "CL2070"). For both lines, the transgenes were introduced into *C. elegans* by gonad microinjection as known in the art and described by Mello *et al.* (Mello *et* EMBO J., 10:3959-3970 [1991]). Marker plasmid pRF4 containing the gene (*rol-6*[su1006]) (SEQ ID NO:9)(pRF4 carries a 4 kb *Eco*RI fragment of *C. elegans* genomic DNA containing the *rol*-6[su1006] collagen gene in the Bluescribe vector [Stratagene]; *See*, Mello *et al.*, *supra*; and Kramer *et al.*, Mol. Cell. Biol., 10:2081-2090 [1990]) was coinjected with the constructs, at approximately 100 ng/μl for each plasmid, into morphologically wild-type animals, and Roller transgenic progeny were recovered. The (*rol-6*[su1006]) gene (SEQ ID NO:9), contained within the pRF4 plasmid is a mutated *C. elegans* collagen gene, the expression of which produces the dominant, distinctive "Roller" phenotype. Figures 9 and 10 show the DNA and amino acid sequences of the *rol-6* gene used in this Example, respectively.

Transmitting lines were established and maintained by selection for the Roller marker phenotype. Transgenic animals produced in this manner maintain the injected DNA as an extrachromosomal, multicopy array of variable mitotic and meiotic stability. Strains containing chromosomally integrated transgenes were recovered by irradiation of lines containing extrachromosomal transgenic arrays with 7000 rad (1 rad = 0.01 Gy) of gamma rays from a Cesium-66 source. Progeny of irradiated animals were then screen for 100% transmittance of the marker transgene. The transgenes in both the CL2005 and CL2070 lines were chromosomally integrated, and were 100% stable.

Transgenic (*i.e.*, as indicated by the Roller phenotype) animals produce both transgenic (*i.e.*, Roller) and non-transgenic (*i.e.*, non-Roller) progeny. These non-transgenic progeny were found to serve as good internal controls for phenotypic and immunohistochemical comparisons. The expression of GFP in the dual transgenic animals can be detected in less than 24 hours after the upshift of animals from 16 to 25.5°C, at all

stages of development, from late embryonic to adult. Expression of GFP resulted in the production of green fluorescence in muscle cells; intense tissue-specific expression of GFP was observed. The fluorescence can be observed using compound or dissecting epifluoroescence microscopy (*i.e.*, with standard fluorescein excitation and emission filters). It is also contemplated that the fluorescence is observable by use of fluorimeters and cell sorters.

EXAMPLE 4

Immunohistochemistry of Transgenic Animals

In this Example, immunohistochemistry was used to confirm the transmittance of the chimeric constructs in large populations of putative integrated lines.

As described by Link (Link, 1995, *supra*), whole mount specimens were prepared by fixing animals in 4% paraformaldehyde and permeabilizing them with 2-mercaptóethanol collagenase as known in the art and described by Link *et al.* (Link *et al.*, Genetics 131:867-881 [1992]). Three antibodies were used (polyclonal rabbit anti-β peptide antibody, Boehringer Mannheim; mouse monoclonal antibody Ab 4G8, available from Senetek; and monoclonal 4.1, a gift from B. Cordell at Scios Nova). The monoclonal 4.1 recognizes residues 8-15 of β-peptide.

As was observed by Link (Link, 1995, *supra*) animals transgenic for the *unc-54*/β(1-42) minigene construct contained muscle-specific deposits of anti-β peptide immunoreactivity. Non anti-β-immunoreactivity was observed in control animals (*i.e.*, non-transgenic, wild-type *C. elegans* tested concurrently). Although *C. elegans* is reported to contain a homolog of the β-amyloid precursor protein gene (*See*, Daigle and Li, Proc. Natl. Acad. Sci., 90:2045-2049 [1993]), this sequence does not contain an apparent β-peptide domain. Thus, it would not be expected, nor did it show, cross-reactivity with the anti-β antibody used in these experiments. For dually transgenic animals, immunoreactive deposits accumulated in the body wall of the animals.

Next, to determine whether the immunoreactive deposits observed in the *unc-54*/β-(1-42) strains displayed the tinctural properties of classic insoluble β-amyloid, transgenic

strains were fixed and stained with thioflavin S, a fluorescent amyloid-specific dye, as known in the art, and described by Guntern and Bouras (R. Guntern and C. Bouras, Experientia 48:-10 [1992]). Thioflavin S-reactive deposits were found in all strains containing the *unc-54*/β-(1-42) minigene constructs, but not in control wild-type animals.

Whole mounts of fixed dual transgenic animals stained with coumarin-phallacidin (a muscle-specific probe) and anti-β peptide antibody. Muscle-specific β peptide deposits were observed, and showed a qualitative correlation with the level of GFP expression in these muscle cells (*i.e.*, cells that contained more β-peptide deposits were more green than cells with less β-peptide deposits).

In addition, Congo Red and Chrysamine G, two dyes known to interact with β-amyloid and have been reported to interfere with its aggregation are tested on the dually transgenic animals. Anti-oxidants (*e.g.*, vitamin E and ascorbate) are also tested. In these experiments, interference with aggregation of β-peptide, or interaction with β-amyloid are observed.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

CLAIMS

What is claimed is:

1. A transparent animal comprising:

(a) a toxic transgene;

(b) an inducible promotor operably linked to a reporter gene, wherein the toxic transgene induces the promotor and wherein the expression of the reporter gene is detectable *in vivo*.

2. The transparent animal of claim 1, wherein the inducible promoter operably linked to the reporter gene is stably integrated into the genome of the animal.

3. The transparent animal of claim 2, wherein the inducible promotor is induced by disruption of cellular function.

4. The transparent animal of claim 2, wherein the inducible promoter is a heat shock promoter.

5. The transparent animal of claim 2, wherein the toxic transgene is β-peptide.

6. The transparent animal of claim 2, wherein said transparent animal is selected from the class Nematoda.

7. The transparent animal of Claim 6, wherein said transparent animal is *Caenorhabditis elegans*.

8. The transparent animal of Claim 2, wherein said genome comprises SEQ ID NO:5 and SEQ ID NO:8.

9. The transparent animal of Claim 2, wherein said reporter gene is green fluorescence protein (GFP).

10. The transparent animal of claim 1, further comprising a heterologous gene marker.

11. A method for producing a dually transgenic non-human animal comprising:

a) providing:

i) a first and second non-human animal; and

ii) a first toxic transgene; and

iii) a second transgene, comprising a reporter;

b) introducing said first toxic transgene into the genome of said first non-human animal to produce a first transgenic animal, and introducing said second transgene into the genome of said second non-human animal to produce a second transgenic animal; and

c) mating said first transgenic animal with said second transgenic animal to produce a dually transgenic animal, wherein said toxic transgene and said reporter are expressed.

12. A dually transgenic non-human animal produced according to the method of claim 11, wherein the toxic transgene is β-peptide.

13. A dually transgenic non-human animal produced according to the method of claim 11, wherein said non-human animal is transparent.

14. The dually transgenic non-human animal of claim 13, wherein said non-human animal is a nematode.

15. The dually transgenic non-human animal of claim 14, wherein said nematode is *Caenorhabditis elegans*

16. The method of claim 11, wherein said reporter is green fluorescent protein.

17. The dually transgenic non-human animal produced according to the method of claim 11, wherein said first transgene comprises pCL25.

18. The dually transgenic non-human animal produced according to the method of claim 11, wherein said second transgene comprises pCL12.

19. The dually transgenic non-human animal produced according to the method of claim 11, further comprising a heterologous marker gene.

20. A method for testing compounds for toxicity, comprising:

a) providing:

i) a dually transgenic non-human animal expressing a toxic transgene and reporter gene operably linked to a promoter inducible by the toxic transgene;

ii) a composition comprising a test compound in a form suitable for administration such that said compound is bioavailable in the cells of said non-human animal; and

b) administering said test compound to said non-human animal.

21. The method according to claim 20, wherein the toxic transgene is β-peptide and the reporter gene is GFP.

22. The method of Claim 21, further comprising c) measuring a reduction in the fluorescence of said non-human animal and thereby identifying a compound as therapeutic.

23. The method of Claim 20, wherein said compounds inactivate said β-peptide expressed by said dually transgenic animal.

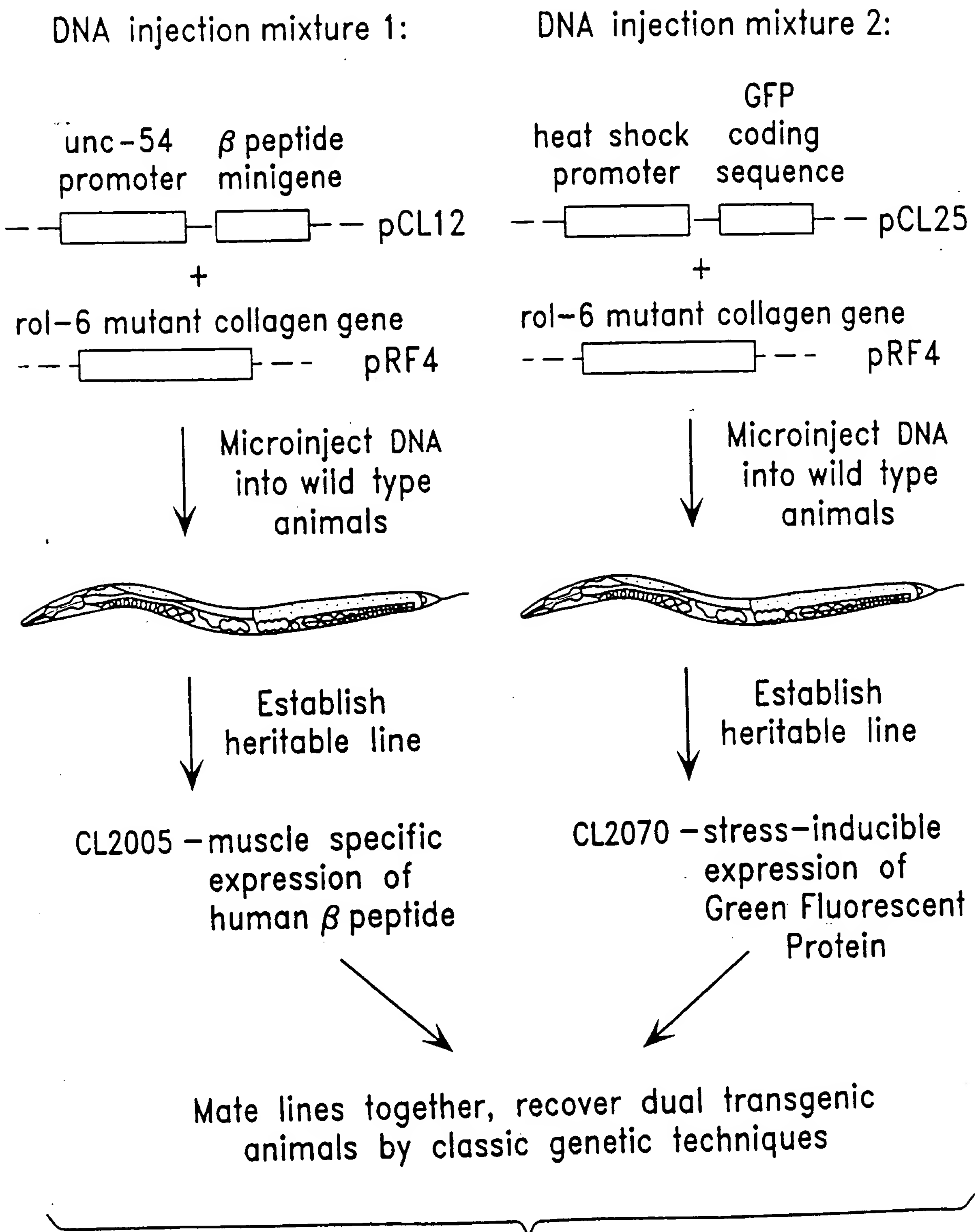

DNA injection mixture 1:
unc-54 promoter
β peptide minigene
pCL12
+
rol-6 mutant collagen gene
pRF4
Microinject DNA into wild type animals
Establish heritable line
CL2005 - muscle specific expression of human β peptide
DNA injection mixture 2:
heat shock promoter
GFP coding sequence
pCL25
+
rol-6 mutant collagen gene
pRF4
Microinject DNA into wild type animals
Establish heritable line
CL2070 - stress-inducible expression of Green Fluorescent Protein
Mate lines together, recover dual transgenic animals by classic genetic techniques

Fig. 1

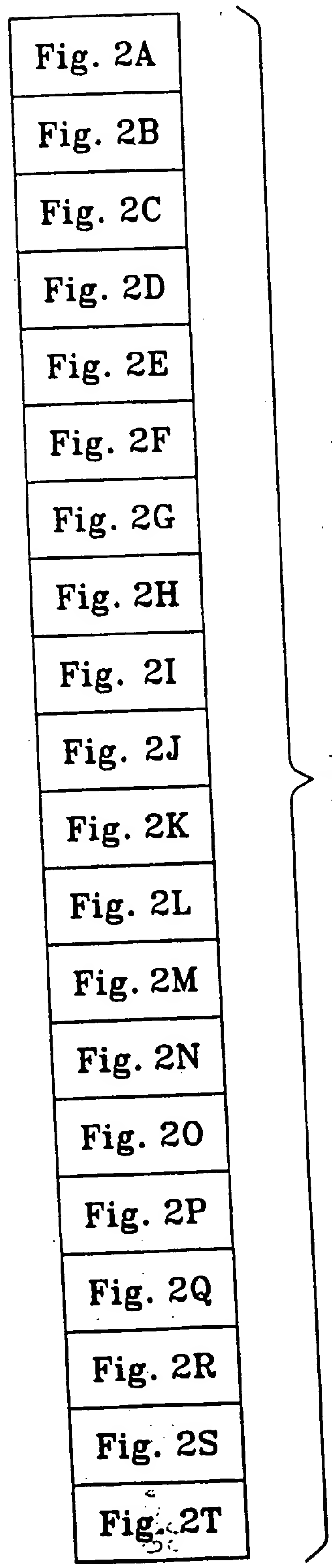


Fig. 2A
Fig. 2B
Fig. 2C
Fig. 2D
Fig. 2E
Fig. 2F
Fig. 2G
Fig. 2H
Fig. 2I
Fig. 2J
Fig. 2K
Fig. 2L
Fig. 2M
Fig. 2N
Fig. 2O
Fig. 2P
Fig. 2Q
Fig. 2R
Fig. 2S
Fig. 2T

Fig. 2

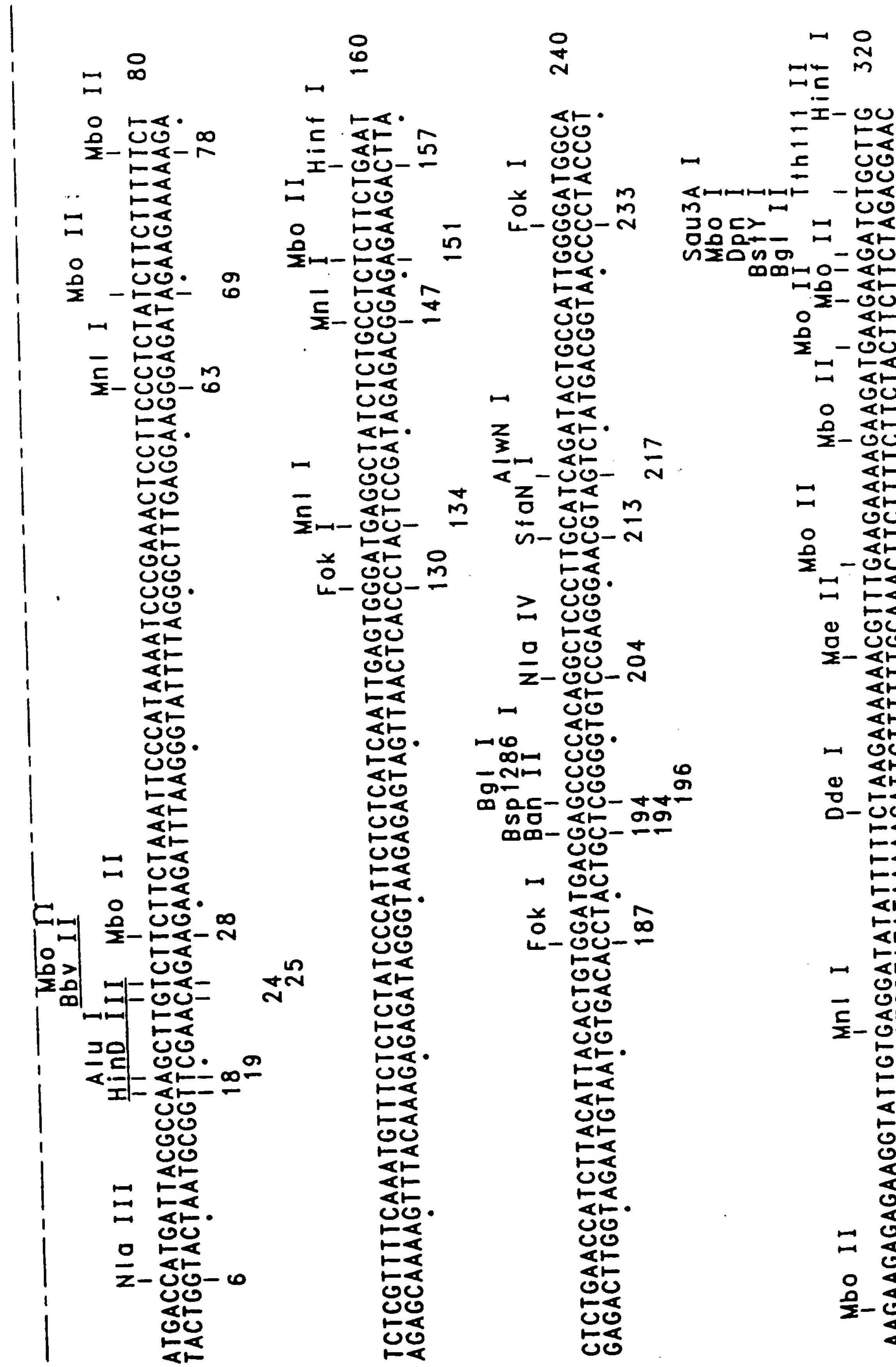

Nla III
Alu I
HinD III
Mbo II
Bbv II
Mbo II
Mnl I
Mbo II
Mbo II
ATGACCATGATTACGCCAAGCTTGTCTTCTTCTAAATTCCCATAAAATCCCGAAACTCCTTCCCTCTATCTTCTTTTTCT 80
TACTGGTACTAATGCGGTTCGAACAGAAGAAGATTTAAGGGTATTTTAGGGCTTTGAGGAAGGGAGATAGAAGAAAAAGA
6
18
19
24
25
28
63
69
78
Fok I
Mnl I
Mnl I
Mbo II
Hinf I
TCTCGTTTTCAAATGTTTCTCTCTATCCCATTCTCTCATCAATTGAGTGGGATGAGGCTATCTCTGCCTCTCTTCTGAAT 160
AGAGCAAAAGTTTACAAAGAGAGATAGGGTAAGAGAGTAGTTAACTCACCCTACTCCGATAGAGACGGAGAGAAGACTTA
130
134
147
151
157
Fok I
Bgl I
Bsp1286 I
Ban II
Nla IV
SfaN I
AlwN I
Fok I
CTCTGAACCATCTTACATTACACTGTGGATGACGAGCCCCACAGGCTCCCTTGCATCAGATACTGCCATTGGGGATGGCA 240
GAGACTTGGTAGAATGTAATGTGACACCTACTGCTCGGGGTGTCCGAGGGAACGTAGTCTATGACGGTAACCCCTACCGT
187
194
194
196
204
213
217
233
Mbo II
Mnl I
Dde I
Mae II
Mbo II
Mbo II
Mbo II
Mbo II
Sau3A I
Mbo I
Dpn I
BstY I
Bgl II
Tth111 II
Hinf I
AAGAAGAGAGAAGGTATTGTGAGGATATATTTTTCTAAGAAAAAACGTTTGAAGAAAAGAAGATGAAGAAGATCTGCTTG 320
TTCTTCTCTCTTCCATAACACTCCTATATAAAAAGATTCTTTTTTGCAAACTTCTTTTCTTCTACTTCTTCTAGACGAAC

Fig. 2A

```
243  261  275  285  291  299  305  308  310  310  311  311  311  315  320

Mae III (341)  Bst B I (358)  Taq I (359)  Mbo II (361)  Mse I (368)  Bsm I (375)  Mse I (395)
ATTCATTGCACAAGTTAGAAGTAACAGGGGTCTATATTTCGAAGAACTTAAAGGGAATGCAACTGAACATAAAATTAAAC  400
TAAGTAACGTGTTCAATCTTCATTGTCCCCAGATATAAAGCTTCTTGAATTTCCCTTACGTTGACTTGTATTTTAATTTG

Hinf I (410)  Pst I (415)  HinC II (476)  Hpa I (476)  Mse I (477)
AAAGGGATTGAATCCTGCAGTGAGTATTTTCGGTTTTTCACTGGTTCTCTGTAAAAAGAGTAATGCAAAGGGCAAGTTAA  480
TTTCCCTAACTTAGGACGTCACTCATAAAAGCCAAAAAGTGACCAAGAGACATTTTTCTCATTACGTTTCCCGTTCAATT

Dde I (481)  Mse I (508)  Mbo II (517)  Bgl II  BstY I  Dpn I  Mbo I  Sau3A I  Mae I
Hph I (528)  Sec I  Mbo II (539)  Mnl I (551)  Mnl I  Mbo II (560)
CTTAGGTCGTAAATGTATTGAATTTGCTTAAAATCTGAAGATCTAGTGGTGAACCGTGGAAGATTATCAAGAGGAGGCTG  560
GAATCCAGCATTTACATAACTTAAACGAATTTTAGACTTCTAGATCACCACTTGGCACCTTCTAATAGTTCTCCTCCGAC
481                        508      517        528        539         551       560
```

Fig. 2B

```
519
519
520
520
520
523
534
554

Sau3A I
Mbo I
Dpn I
BstY I
Bgl II     Mse I      Mse I
                      Ase I
AAGATCTGTTTAAGAACCATTAATCAAACTGGTATTCTATTTTCACTGGTTGTATGTAAACATTCTATCTTATTCCTTTT  640
TTCTAGACAAATTCTTGGTAATTAGTTTGACCATAAGATAAAAGTGACCAACATACATTTGTAAGATAGAATAAGGAAAA
562        570        579
562                    580
563
563
563

                                                                        Sau3A I
                                                                        Mbo I
                               HinC II    Rsa I      EcoR I             Dpn I
ATCACTGTTCTGCACTTTCCTATAAAAAAAGTTGACCGACCGTACTCTCTGAATTCATTTTTCCCGATCTTACCAACTCC  720
TAGTGACAAGACGTGAAAGGATATTTTTTTCAACTGGCTGGCATGAGAGACTTAAGTAAAAAGGGCTAGAATGGTTGAGG
                               671        682        691                706
                                                                        706
                                                                        706

              ScrF I                                                        Mnl I
Sau3A I       EcoR II          HgiA I                                 Nla IV
Mbo I         BstN I           Bsp1286 I                    Mbo II    Ban I
Dpn I         Sec I       Mbo II          EcoR I
CGATCTATCTCTATCCCTGGTTTTTTCTTCGTGCTCCAATGGAATTCTTGAGACTTCCACTATCTTCTCTGGCACCCTCC  800
GCTAGATAGAGATAGGGACCAAAAAAGAAGCACGAGGTTACCTTAAGAACTCTGAAGGTGATAGAAGAGACCGTGGGAGG
722           735         746             762                   783       791
722            736              751                                       791
```

Fig. 2C

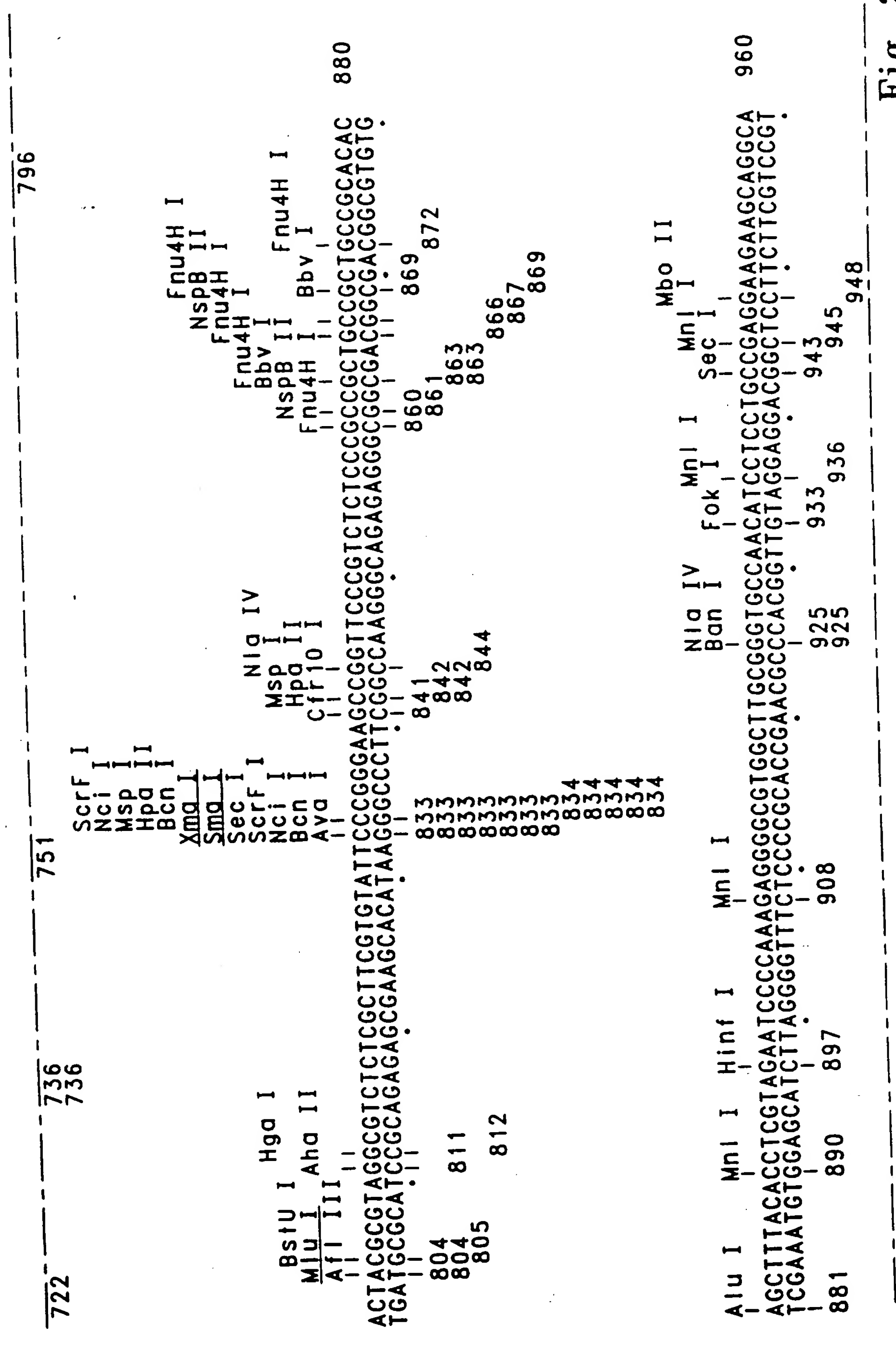


Fig. 2D

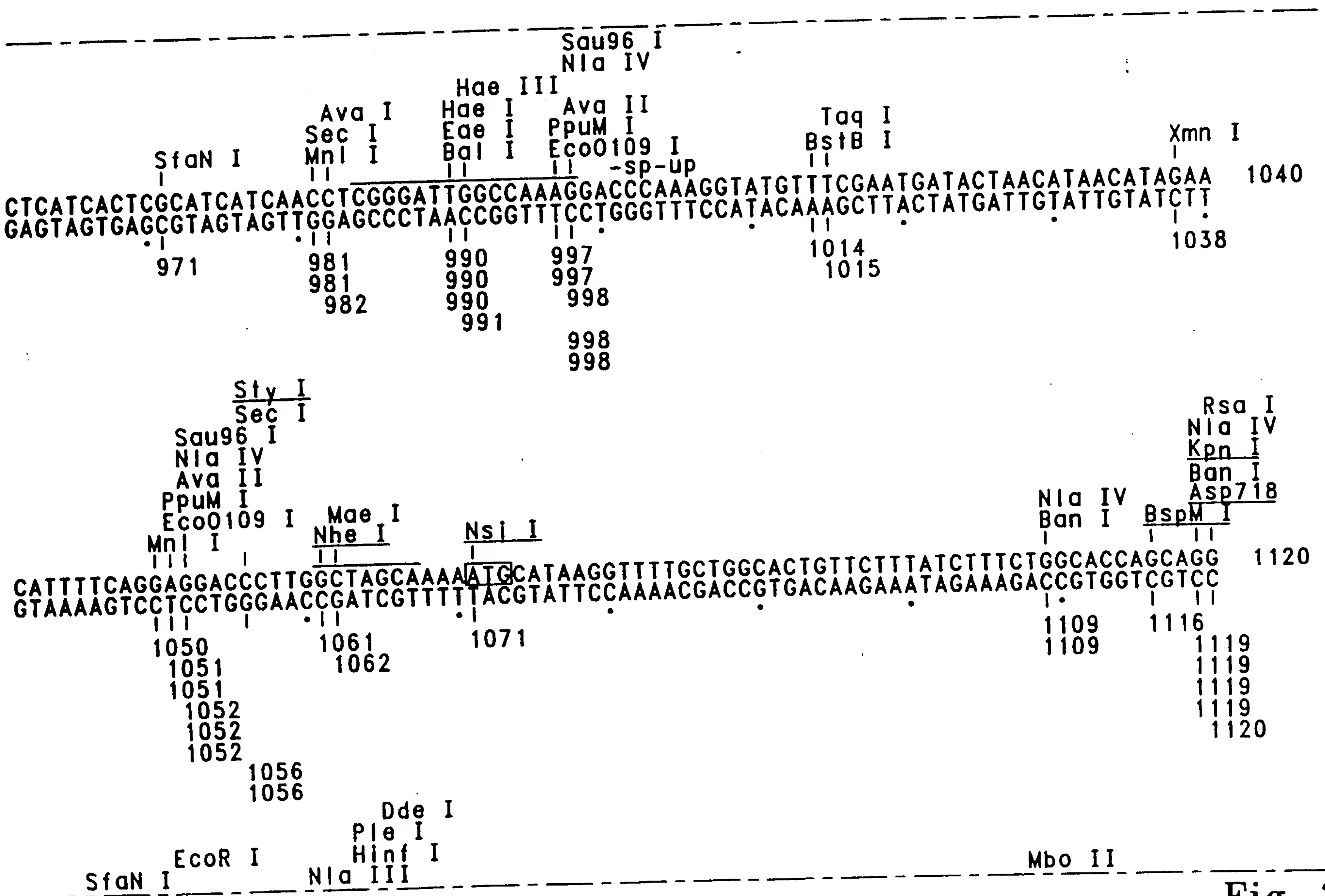

Sau96 I
Nla IV
Hae III
Ava I
Hae I
Ava II
Sec I
Eae I
PpuM I
Taq I
SfaN I
Mnl I
Bal I
EcoO109 I
BstB I
Xmn I
-sp-up
CTCATCACTCGCATCATCAACCTCGGGATTGGCCAAAGGACCCAAAGGTATGTTTCGAATGATACTAACATAACATAGAA
GAGTAGTGAGCGTAGTAGTTGGAGCCCTAACCGGTTTCCTGGGTTTCCATACAAAGCTTACTATGATTGTATTGTATCTT
1040
1038
971
981
981
982
990
990
990
991
997
997
998
998
998
1014
1015
Sty I
Sec I
Sau96 I
Nla IV
Ava II
PpuM I
EcoO109 I
Mnl I
Mae I
Nhe I
Nsi I
Nla IV
Ban I
BspM I
Rsa I
Nla IV
Kpn I
Ban I
Asp718
CATTTTCAGGAGGACCCTTGGCTAGCAAAAATGCATAAGGTTTTGCTGGCACTGTTCTTTATCTTTCTGGCACCAGCAGG
GTAAAAGTCCTCCTGGGAACCGATCGTTTTTACGTATTCCAAAACGACCGTGACAAGAAATAGAAAGACCGTGGTCGTCC
1120
1050
1051
1051
1052
1052
1052
1056
1056
1061
1062
1071
1109
1109
1116
1119
1119
1119
1119
1120
Dde I
Ple I
Hinf I
EcoR I
Nla III
SfaN I
Mbo II

Fig. 2E

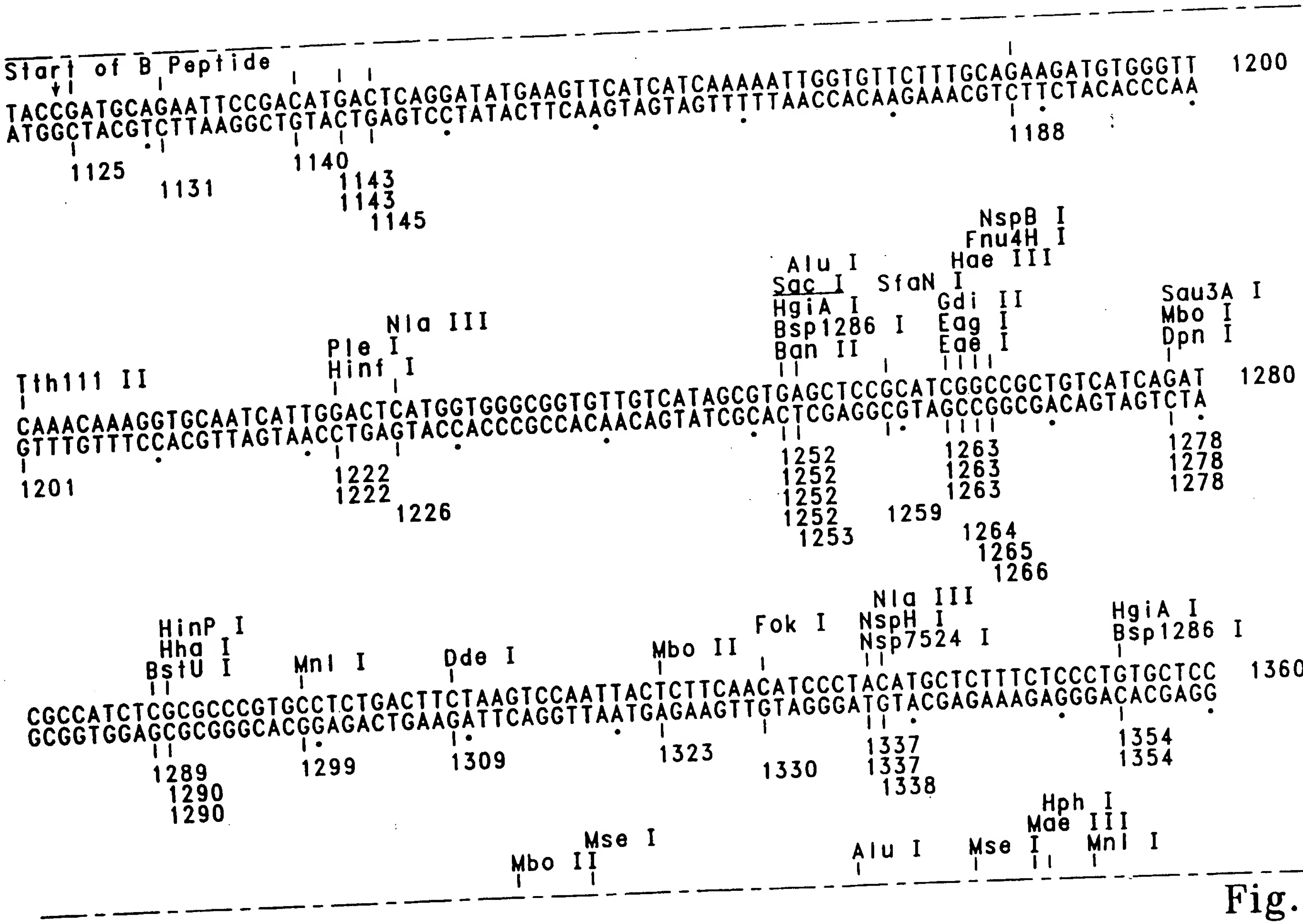

Start of B Peptide
TACCGATGCAGAATTCCGACATGACTCAGGATATGAAGTTCATCATCAAAAATTGGTGTTCTTTGCAGAAGATGTGGGTT 1200
ATGGCTACGTCTTAAGGCTGTACTGAGTCCTATACTTCAAGTAGTAGTTTTTAACCACAAGAAACGTCTTCTACACCCAA
1125 1131 1140 1143 1143 1145 1188
Tth111 II
Ple I
Hinf I
Nla III
Alu I
Sac I
HgiA I
Bsp1286 I
Ban II
SfaN I
NspB I
Fnu4H I
Hae III
Gdi II
Eag I
Eae I
Sau3A I
Mbo I
Dpn I
CAAACAAAGGTGCAATCATTGGACTCATGGTGGGCGGTGTTGTCATAGCGTGAGCTCCGCATCGGCCGCTGTCATCAGAT 1280
GTTTGTTTCCACGTTAGTAACCTGAGTACCACCCGCCACAACAGTATCGCACTCGAGGCGTAGCCGGCGACAGTAGTCTA
1201 1222 1222 1226 1252 1252 1252 1252 1253 1259 1263 1263 1263 1264 1265 1266 1278 1278 1278
HinP I
Hha I
BstU I
Mnl I
Dde I
Mbo II
Fok I
Nla III
NspH I
Nsp7524 I
HgiA I
Bsp1286 I
CGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCC 1360
GCGGTGGAGCGCGGGCACGGAGACTGAAGATTCAGGTTAATGAGAAGTTGTAGGGATGTACGAGAAAGAGGGACACGAGG
1289 1290 1290 1299 1309 1323 1330 1337 1337 1338 1354 1354
Mbo II
Mse I
Alu I
Mse I
Hph I
Mae III
Mnl I

Fig. 2F

```
CACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACA  1440
GTGGGGGATAAAAACAATAATAGTTTTTTTGAAGAAGAATTAAAGAAACAAAAAATCGAAGAAAATTCAGTGGAGATTGT
1393 1398 1416 1424 1428 1429 1432

Hinf I; Mse I, Ase I, Xmn I; Taq I; HgiA I, Bsp1286 I, Mnl I; Mse I, Ase I
ATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATA  1520
TACTTTAACACATCTAAGTTTTTATCTTAATTAAGCATTATTTTTCAGCTTTTTTTAACACGAGGGAGGGGGGTAATTAT
1454 1466 1468 1469 1487 1499 1499 1505 1514 1515

Rsa I
ATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTACTTCTGATAAATTTTTTTTGAAA  1600
TATTAAGATAGGGTTTTAGATGTGTTACAAGACACATGTGAAGAATACAAAAAAATGAAGACTATTTAAAAAAAACTTT
1555

Nde I; Mae III; Mae II; Mbo II; Mae II
CATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCA  1680
GTAGTATCTTTTTTGGCGTGTGTTTTATGGAATAGTATACAATGCAAAGTCAAATACTGGCGTTAAAAATAAAGAAGCGT
1635 1640 1643 1673 1680

Sau96 I, Hae III, Mnl I, BspH I, Nla III; Aat II, Aha II, Mae II; Nla III
CGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGA  1760
GCAGACCCGGAGAGTACTGCAGTTTAGTACGAGTAGCACTTTTTCAAAACCTCATAAAAACCTTAAAAAGTTAGTTCACT
1686 1697 1707
```

Fig. 2G

```
1687 1689 1693 1694 1697 1698
Ase I  Mse I          Tth111 II        Taq I  Mnl I
AAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTT  1840
TTCAAATACTTTAATTAAAAGGACGAAAACGAAAAACCCCCAAAGGGGATAACAAACAGTTCTCAAAGCTCCTGCCGCAA
1772 1773                                  1812             1827 1829

HgiA I  Bsp1286 I  SfaN I  Dpn I  Mbo I  Sau3A I  Mbo II  Mnl I  Dde I
TTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAA  1920
AAAGAACGATTTTAGTGTTCATAACTACTCGTGCTACGTTCTTTCTAGCCTTCTTCCAAACCCAAACTCCGAGTCACCTT
1868 1868 1874 1885 1885 1885 1890 1907 1911

Hph I                                             Mse I      Xmn I
GGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATA  2000
CCACTCATCTTCAACTATTAAACTTTCACCTCATCACAGATACCCCAAAAACGGAATTTACTGTCTTATGTAAGGGTTAT
1921                                              1975       1985

Tth111 I  Dra I  Mse I  Mse I                     Dra I  Mse I          Mse I
TACCAAACATAACTGTTTAAAATTAAACATTTTTCTAAATTTTATATGATTTCTTTTAAATTTGCAAAAATTACTTAAAT  2080
ATGGTTTGTATTGACAAATTTTAATTTGTAAAAAGATTTAAAATATACTAAAGAAAATTTAAACGTTTTTAATGAATTTA
2004  2016 2017  2023                            2055 2056              2075

BstU I  HinP I                           Msp I   Mse I
```

Fig. 2H

```
 EcoR I  Hha I      Mae III                          Hpa II  Ase I         Tth111 II
TTGAATTCCCGCGCAAATGAGTGACTTCATTTTCTGCATTATTGTGTTTTCCGGCTATATTAATAGGTATTTGTTTGTGT   2160
AACTTAAGGGCGCGTTTACTCACTGAAGTAAAAGACGTAATAACACAAAAGGCCGATATAATTATCCATAAACAAACACA
 2083    2091     2101                               2131    2139          2152
        2090                                         2131     2140
         2091

              Taq I
             BstB I               Taq I                                  Mse I
           Hinf I                BstB I                                 Ase I
TTTTCTTTATTTTATGATTCGAACTCCAATTTGTAAATTTTCGAACATATTTCCCTAAAGAAAAAAATATGATTAATCTGG   2240
AAAAGAAATAAAATACTAAGCTTGAGGTTAAACATTTAAAAGCTTGTATAAAGGGATTTCTTTTTTTATACTAATTAGACC
              2176                2200                                  2231
               2178                2201                                  2232
                2179

                                                                     Hae III
                                                                    Hae I
                                                                    Eae I
                                                 Mbo II             Bal I
AAAAATTGGAAAATTATTTTTCAAATAAAAAACAAAGAAAAAAATGAAGAAAAACCTATTAGTTTGGCCATAAAACGCAA   2320
TTTTTAACCTTTTAATAAAAAGTTTATTTTTTGTTTCTTTTTTTACTTCTTTTTGGATAATCAAACCGGTATTTTGCGTT
                                                 2286               2305
                                                                    2305
                                                                    2305
                                                                     2306

               Mae III
              Mae I             BstU I
             Aha I             HinP I
    Taq I    Aat I             Hha I                    Hae III
AAATGTCGAAAATGACGTCACTCATCTGCGCGGGAATCAAGAATAATTCGGCCTTTTTTATTTTTTTGGAAAATCGTAA   2400
TTTACAGCTTTTACTGCAGTGAGTAGACGCGCCCTTTAGTTCTTATTAAGCCGGAAAAAATAAAAAAACCTTTTAGCATT
    2326     2334              2348                     2371
             2334              2348
```

Fig. 2I

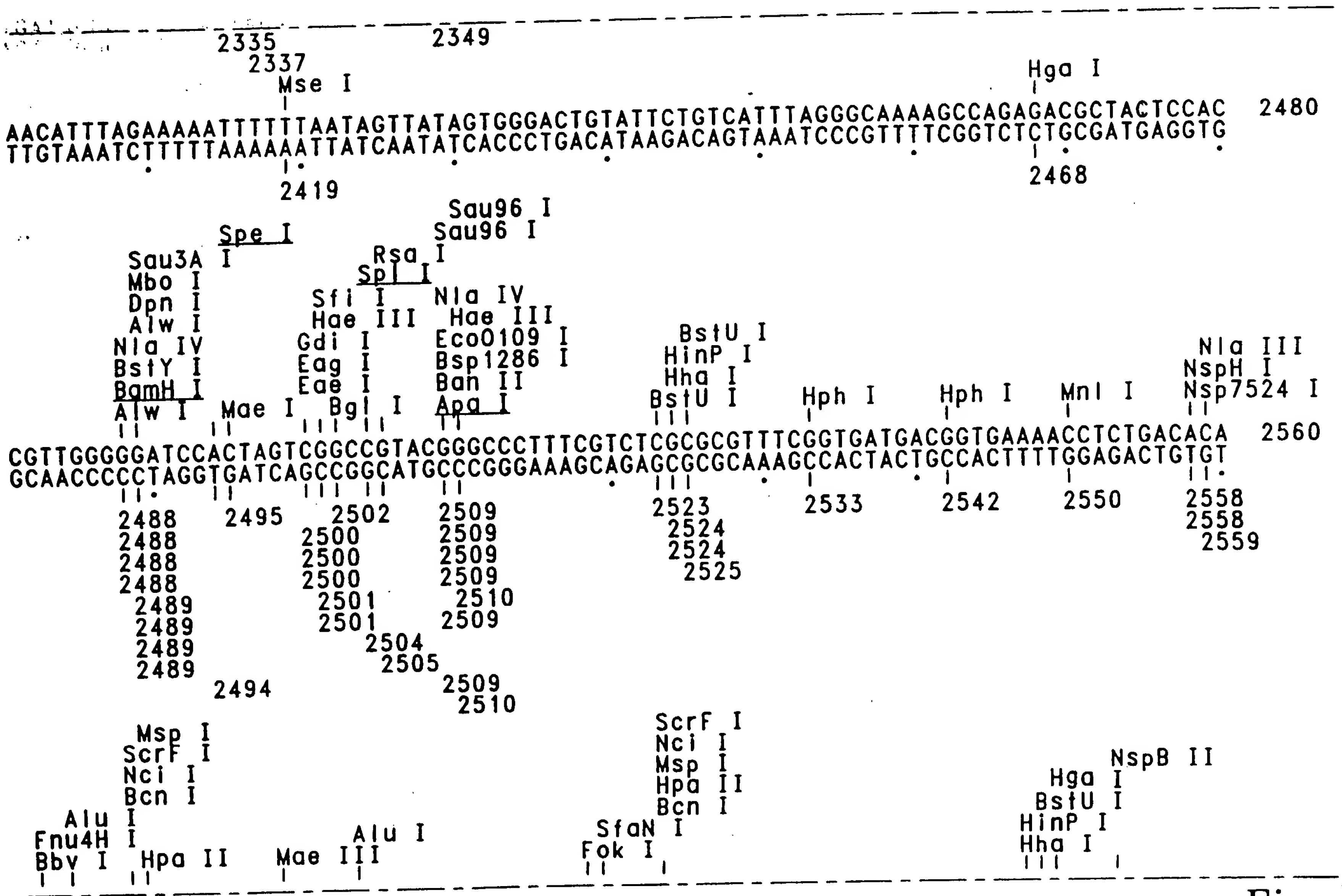

2335
2337
Mse I
2349
Hga I
AACATTTAGAAAAATTTTTTAATAGTTATAGTGGGACTGTATTCTGTCATTTAGGGCAAAAGCCAGAGACGCTACTCCAC
TTGTAAATCTTTTTAAAAAATTATCAATATCACCCTGACATAAGACAGTAAATCCCGTTTTCGGTCTCTGCGATGAGGTG
2480
2419
2468
Sau3A I
Mbo I
Dpn I
Alw I
Nla IV
BstY I
BamH I
Alw I
Spe I
Mae I
Sfi I
Hae III
Gdi I
Eag I
Eae I
Bgl I
Rsa I
Spl I
Sau96 I
Sau96 I
Nla IV
Hae III
Eco0109 I
Bsp1286 I
Ban II
Apa I
BstU I
HinP I
Hha I
BstU I
Hph I
Hph I
Mnl I
Nla III
NspH I
Nsp7524 I
CGTTGGGGATCCACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACA
GCAACCCCTAGGTGATCAGCCGGCATGCCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGT
2560
2488
2488
2488
2488
2489
2489
2489
2489
2494
2495
2500
2500
2500
2501
2501
2502
2504
2505
2509
2509
2509
2509
2510
2509
2509
2510
2523
2524
2524
2525
2533
2542
2550
2558
2558
2559
Msp I
ScrF I
Nci I
Bcn I
Alu I
Fnu4H I
Bby I
Hpa II
Mae III
Alu I
ScrF I
Nci I
Msp I
Hpa II
Bcn I
SfaN I
Fok I
NspB II
Hga I
BstU I
HinP I
Hha I

Fig. 2J

```
TGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGT  2640
ACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCA
2562 2569 2578 2583 2598 2599 2603 2627 2627 2628 2629 2633
2562
2564
2568
2568
2568
2569
2603
2603
2603
2603

                                          HgiA I
                                          Bsp1286 I
                 SfaN I           Dde I        Nde I
      Mse I  Fnu4H I          Rsa I      ApaL I
GTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAAT  2720
CAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTA
      2663   2671 2674          2689 2692  2698  2704
                                           2698
                                           2698

                           Mse I
                           Afl II
                          Bgl I      Mnl I
                         Sfi I     Sau96 I
                       Fnu4H I   Eco0109 I
   SfaN I         SfaN I  Hae III  Hae III                          Mse I
ACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAAT  2800
TGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCCGGAATTCCCGGAGCACTATGCGGATAAAAATATCCAATTA
   2729           2750  2757 2759 2759 2760 2762 2763 2766 2767 2768 2770      2796

            Mae II                      Nla IV
            Aha II                      BstU I
```

Fig. 2K

```
 NlaIII   BspH I (2802, 2803); Dde I (2820); Aat I (2824, 2824, 2825); HinP I, Hha I (2854, 2854, 2855, 2858)
GTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTT  2880
CAGTACTATTATTACCAAAGAATCTGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAA

 Nla III, BspH I (2907, 2908); Ssp I (2940); Mbo II (2953)
CTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA  2960
GATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAT

 Fnu4H I (3002); Hph I (3028)
TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACG  3040
ACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGC

 Hph I (3043); SfaN I (3056); Mbo II (3062); Sau3A I, Mbo I, Dpn I (3065, 3065, 3065);
 HgiA I, Bsp1286 I, ApaL I (3075, 3075, 3075); Mae III (3087); Taq I (3093);
 Sau3A I, Mbo I, Dpn I, BstY I, Alw I (3100, 3100, 3101, 3101, 3101); NspB II (3109);
 Sau3A I, Mbo I, Dpn I, BstY I, Alw I (3118, 3117, 3118, 3118, 3118)
CTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGAT  3120
GACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTA

 Mae II, Xmn I;  HgiA I;  Dra I, Mse I;  BstU I, HinP I
```

Fig. 2L

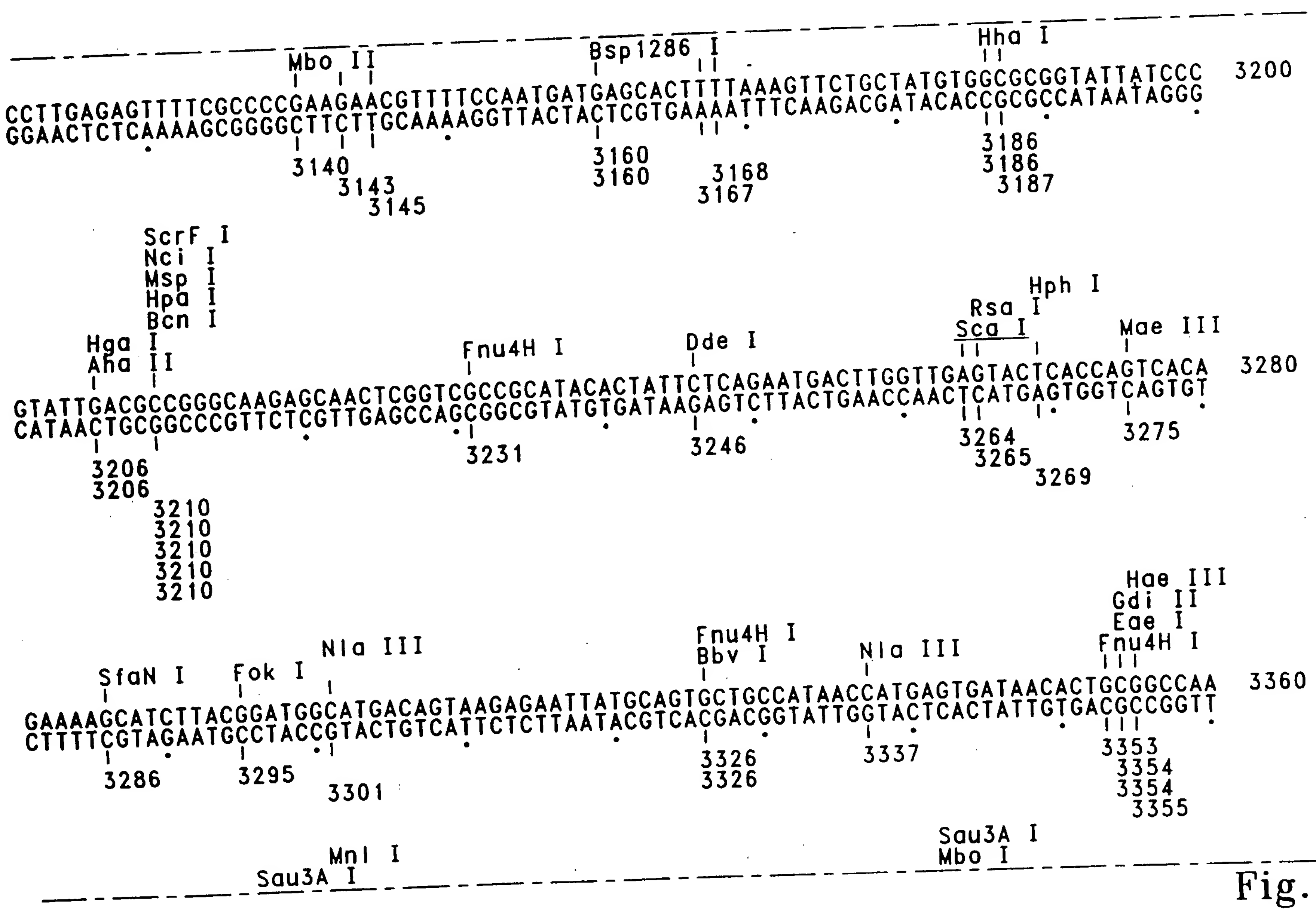

Mbo II
Bsp1286 I
Hha I
CCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCC 3200
GGAACTCTCAAAAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGG
3140 3143 3145 3160 3160 3168 3167 3186 3186 3187
ScrF I
Nci I
Msp I
Hpa I
Bcn I
Hga I
Aha II
Fnu4H I
Dde I
Rsa I
Sca I
Hph I
Mae III
GTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACA 3280
CATAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGT
3206 3206 3210 3210 3210 3210 3210 3231 3246 3264 3265 3269 3275
SfaN I
Fok I
Nla III
Fnu4H I
Bbv I
Nla III
Hae III
Gdi II
Eae I
Fnu4H I
GAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA 3360
CTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTT
3286 3295 3301 3326 3326 3337 3353 3354 3354 3355
Mnl I
Sau3A I
Sau3A I
Mbo I

Fig. 2M

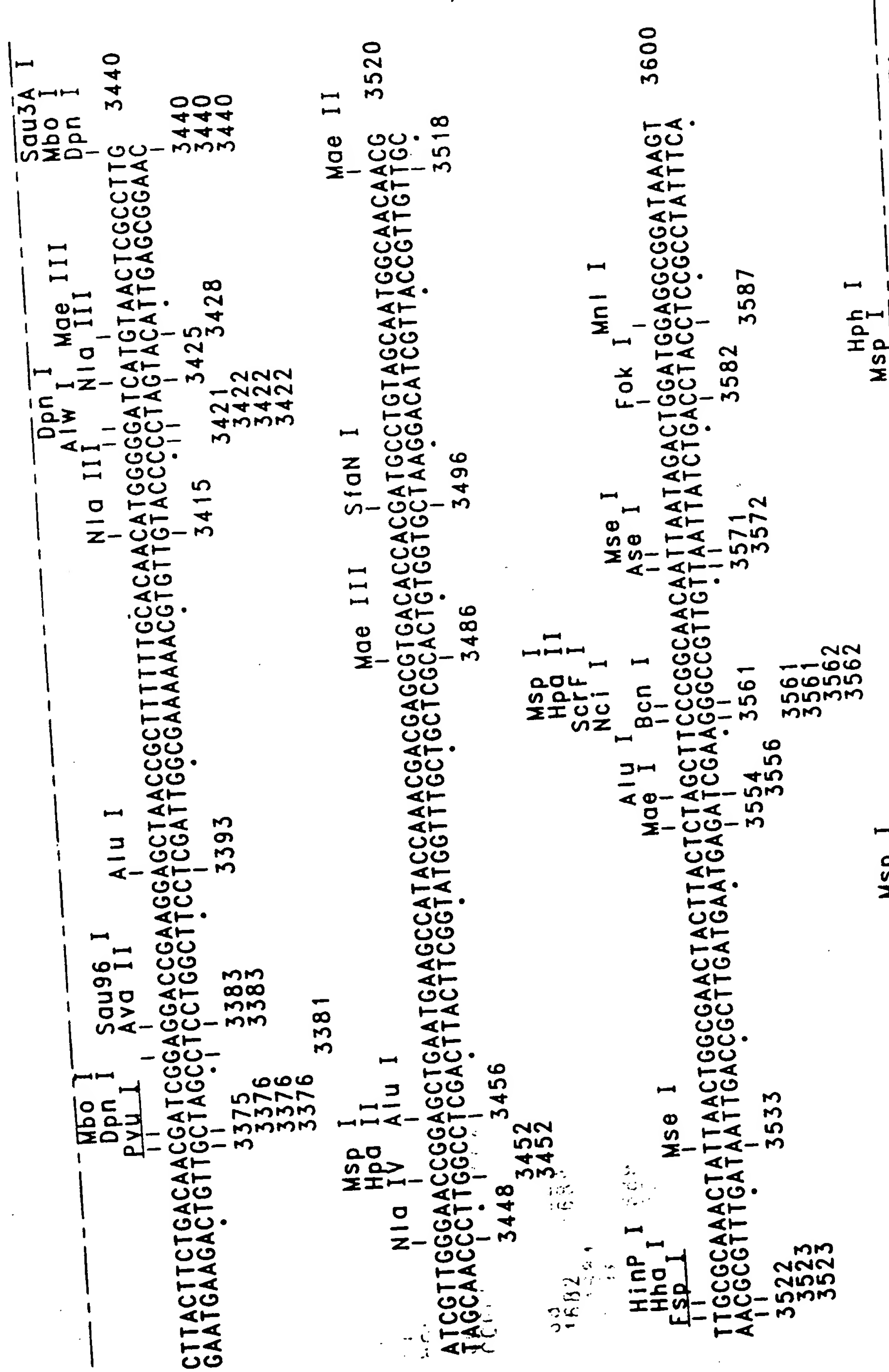

Mbo I
Dpn I
Pvu I
Sau96 I
Ava II
Alu I
Nla III
Alw I
Dpn I
Nla III
Mae III
Sau3A I
Mbo I
Dpn I
CTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTG 3440
GAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAAC
3375 3376 3376 3376 3381 3383 3383 3393 3415 3421 3422 3422 3422 3425 3428 3440 3440 3440
Nla IV
Msp I
Hpa II
Alu I
Mae III
SfaN I
Mae II
ATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACG 3520
TAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTAAGGACATCGTTACCGTTGTTGC
3448 3452 3452 3456 3486 3496 3518
HinP I
Hha I
Esp I
Mse I
Mae I
Alu I
Nci I
ScrF I
Hpa II
Msp I
Bcn I
Ase I
Mse I
Fok I
Mnl I
TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGT 3600
AACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCA
3522 3523 3523 3533 3554 3556 3561 3561 3561 3562 3562 3571 3572 3582 3587
Sau96 I
Msp I
Msp I
Hph I

Fig. 2N

```
     Sau96 I     HinP I  Hae III                                    Hpa II
     Ava II      Hha I   Bgl I  Hpa II                       Nla IV Cfr10 I             BstU I
TGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTC  3680
ACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAG
     3605        3616   3623   3629                          3659                     3680
     3605        3616  3622                                    3662
                       3622                                     3663
                               3629                             3663
                                                                  3665

                 Hae III
        Fnu4H I  Sau96 I                                                 Ple I
        Bbv I   Nla IV          Mnl I                                    Hinf I
GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACT  3760
CGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGA
        3692    3700            3717                                     3748
        3692     3701                                                    3748
                  3702
                     Dde I
                Sau3A I        Mnl I
                Mbo I       Nla IV
   Fok I        Dpn I       Ban I        Mse I      Mae III
ATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTC  3840
TACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAG
   3763         3781        3794         3807       3817
                3781        3794
                3781          3798
                    3786

                                                    Sau3A I
                                     Mae I          Mbo I
                                   Sau3A I          Dpn I
                                   Mbo I            Alw I
                         Mse I     Dpn I            BstY I
               Mse I     Dra I     BstY I        Mbo II                     Nla III
```

Fig. 20

```
Dra I   Mse I   Alw I   Hph I   BspH I
ATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGA   3920
TATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACT
3859 3860 3874 3878 3879 3885 3885 3886 3886 3886 3889 3892 3895 3897 3898 3898 3898 3898 3915 3916

Mse I   Mae II   Dde I   Hga I   Sau3A I   Mbo I   Dpn I   Sau3A I   Mbo I   Dpn I   BstY I   Alw I   Mbo II   Sau3A I   Mbo I   Dpn I   Alw I   BstY I
CCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT   4000
GGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGA
3931 3934 3952 3956 3976 3976 3976 3983 3983 3986 3984 3984 3984 3994 3995 3995 3995 3995

BstU I   HinP I   Hha I   Fnu4H I   Bbv I   Tth111 II   Tth111 II   NspB II   Tth111 II   Msp I   Hpa II   Alw I   Dpn I   Mbo I   Sau3A I   Alu I
TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCT   4080
AAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGA
4009 4020 4028 4050 4061 4077
```

Fig. 2P

```
4009 4010   4020 4022   4067 4067 4069 4070 4070 4070

           Mae III          HinP I  Hha I              Mae I            Hae III  Hae I
ACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG   4160
TGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATC
           4100             4118 4118                  4142             4159 4160

                                  Mnl I            Mae III  AlwN I  Fnu4H I  Bbv I  Fnu4H I  Bbv I
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC   4240
CGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCG
                                  4198             4216  4221  4226 4229 4226 4229

       ScrF I  Nci I  Msp I  Hpa II  Bcn I   Ple I  Hinf I   Mae III   Msp I  Hpa II   HinP I  Hha I  NspB II  Fnu4H I  Bbv I
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC   4320
CTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAG
       4257 4257 4257 4257 4257   4265 4265   4279   4283 4283   4292 4292 4294 4294 4295
```

Fig. 2Q

```
HgiA I
Bsp1286 I                                          HinP I
ApaL I      Alu I               Dde I              Hha I
                                                   Hae II
GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGC  4400
CACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGTAACTCTTTCGCGGTGCG
4321        4334                4361                                       4391
4321                                                                        4392
4321                                                                        4392

                                                                          Sec I
                                                                          ScrF I
                                                         Mnl I            EcoR II
                        Msp I    Fnu4H I            HinP I                BstN I
                        Hpa II                      Hha I         Alu I
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG  4480
AAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCC
                        4430     4437               4459              4470 4475
                        4430                        4459     4465          4475
                                                                           4475
                                                                           4475

     ScrF I                                            Taq I
     EcoR II                        Mnl I          Hga I        SfaN I
     BstN I
GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG  4560
CCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCC
     4488                           4522           4534         4548
     4488                                             4537
     4488

                                               Hae III
                                               Hae I
                                               ScrF I                  Nla III
                         Hae III               EcoR II                 NspH I
                         Fnu4H I               BstN I      Hae III     Nsp7524 I
     Nla IV              BstU I        Nla IV              Hae I       Afl III
```

Fig. 2R

```
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT  4640
CCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACA
4566  4591 4592 4594  4605  4609 4609 4609 4611 4612  4622 4623  4635 4635 4635 4636

Hinf I                                Alu I          Fnu4H I  Bbv I  Fnu4H I
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA  4720
AGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCT
4661                                  4696           4710 4713 4713

Ple I  Fnu4H I  Bbv I  HinP I  Hha I  Hinf I  Mnl I  Mbo II  HinP I  Hha I  Hae II  Mnl I  Gdi II  Eae I  BstU I  HinP I  Hha I  BstU I  Hae III
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCC  4800
TGCTGGCTCGCGTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGG
4729 4729 4731 4731 4736 4736  4748  4757  4761 4762 4762  4781  4789 4790 4790 4791  4797 4796 4796
```

Pvu II
Fnu4H I
Mse I Alu I
Ase I NspB II HinP I Mse I

Fig. 2S

```
Hinf I         Bbv I                                                          Hha I         Ase I         Alu I
|     ||        |||                                                            |             ||            |
GATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGC  4880
CTAAGTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCG
|     ||        |||                                                            |             ||            |
                4812                                                           4855          4865          4878
                 4813                                                          4855           4866
4801              4814
     4806      4812
      4807      4813
                   ScrF I
                   EcoR II
                   BstN I
                  Sec I
          Nla IV                            Msp I
          Ban I                             Hpa II
          |    ||                           |
TCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT  4960
AGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAA
          |    ||                           |
          4891                              4919
          4891                              4919
               4896
                4897
                4897
                4897
             Alu I
             |
CACACAGGAAACAGCT  4976
GTGTGTCCTTTGTCGA
             |
             4973
```

Fig. 2T

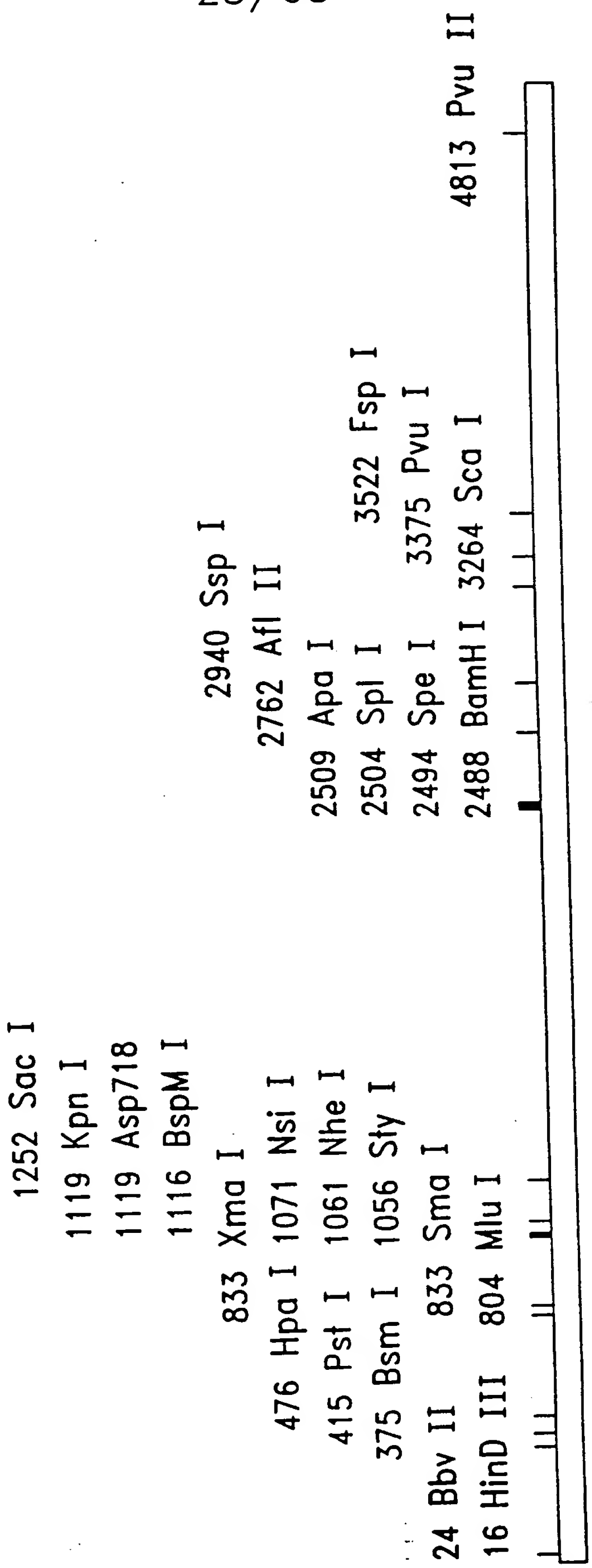

16 HinD III
24 Bbv II
375 Bsm I
415 Pst I
476 Hpa I
804 Mlu I
833 Sma I
833 Xma I
1056 Sty I
1061 Nhe I
1071 Nsi I
1116 BspM I
1119 Asp718
1119 Kpn I
1252 Sac I
2488 BamH I
2494 Spe I
2504 Spl I
2509 Apa I
2762 Afl II
2940 Ssp I
3264 Sca I
3375 Pvu I
3522 Fsp I
4813 Pvu II

Fig. 3

```
1071 /    1                                       1101 /    11
ATG CAT AAG GTT TTG CTG GCA CTG TTC TTT ATC TTT CTG GCA CCA GCA GGT ACC GAT GCA
met his lys val leu leu ala leu phe phe ile phe leu ala pro ala gly thr asp ala
1131 /    21                                      1161 /    31
GAA TTC CGA CAT GAC TCA GGA TAT GAA GTT CAT CAT CAA AAA TTG GTG TTC TTT GCA GAA
glu phe arg his asp ser gly tyr glu val his his gln lys leu val phe phe ala glu
1191 /    41                                      1221 /    51
GAT GTG GGT TCA AAC AAA GGT GCA ATC ATT GGA CTC ATG GTG GGC GGT GTT GTC ATA GCG
asp val gly ser asn lys gly ala ile ile gly leu met val gly gly val val ile ala
1251 /    61
TGA
OPA
```

Fig. 4

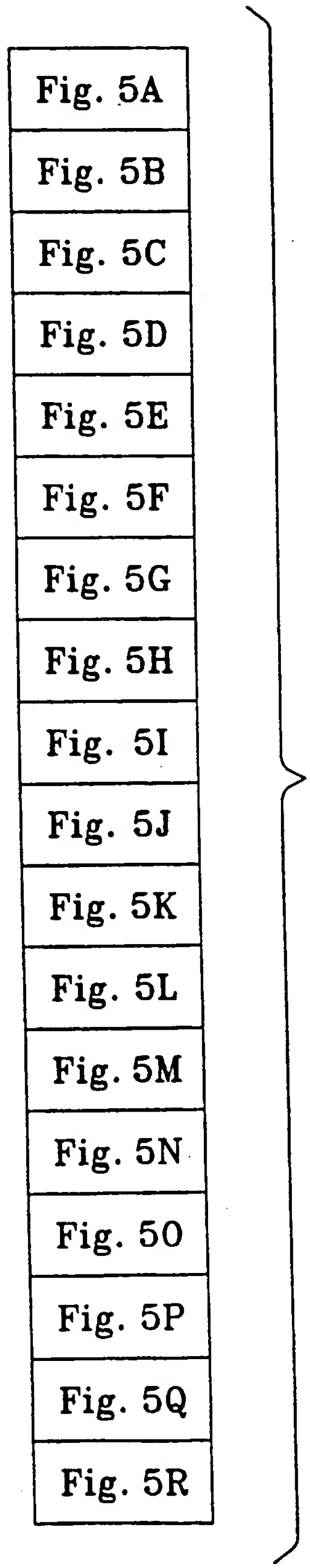


Fig. 5A
Fig. 5B
Fig. 5C
Fig. 5D
Fig. 5E
Fig. 5F
Fig. 5G
Fig. 5H
Fig. 5I
Fig. 5J
Fig. 5K
Fig. 5L
Fig. 5M
Fig. 5N
Fig. 5O
Fig. 5P
Fig. 5Q
Fig. 5R

Fig. 5

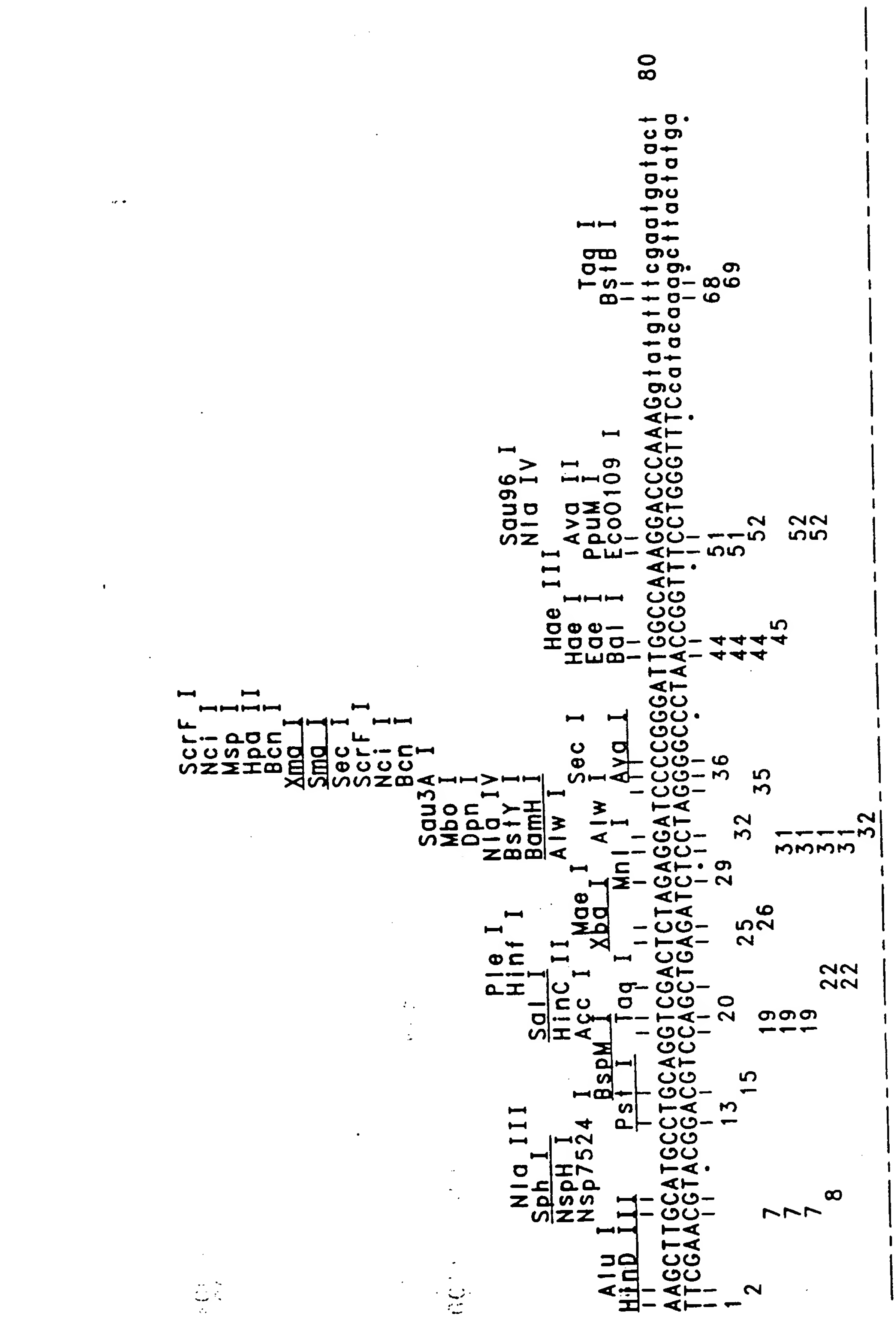

AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGATTGGCCAAAGGACCCAAAGgtatgtttcgaatgatact
TTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCCTAACCGGTTTCCTGGGTTTCcatacaaagcttactatga

Fig. 5A

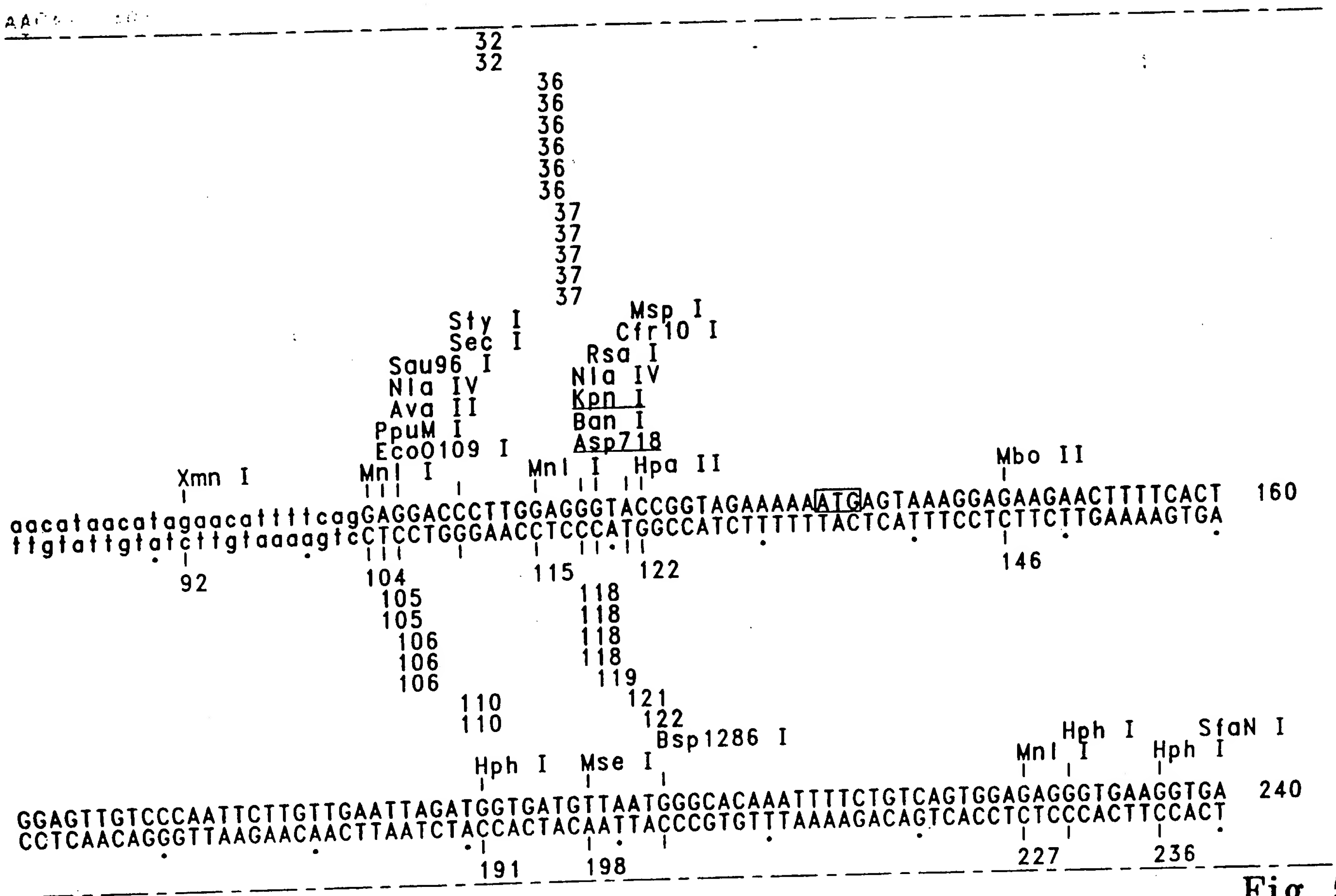

32
36
37
Sty I
Sec I
Sau96 I
Nla IV
Ava II
PpuM I
EcoO109 I
Xmn I
Mnl I
Msp I
Cfr10 I
Rsa I
Nla IV
Kpn I
Ban I
Asp718
Mnl I
Hpa II
Mbo II
aacataacatagaacattttcagGAGGACCCTTGGAGGGTACCGGTAGAAAAAATGAGTAAAGGAGAAGAACTTTTCACT 160
ttgtattgtatcttgtaaaagtcCTCCTGGGAACCTCCCATGGCCATCTTTTTTACTCATTTCCTCTTCTTGAAAAGTGA
92
104
105
106
110
115
118
119
121
122
146
Bsp1286 I
Hph I
Mse I
Mnl I
Hph I
Hph I
SfaN I
GGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGA 240
CCTCAACAGGGTTAAGAACAACTTAATCTACCACTACAATTACCCGTGTTTAAAAGACAGTCACCTCTCCCACTTCCACT
191
198
227
236

Fig. 5B

```
                                          203                          230     239
                                                                 Hae III
                                                                 Hae I
                                                                 Eae I
                                                                 Bal I
                                                               Nla III
                                                              Sty I
                                                              Sec I
                     Mse I                                    Nco I              Mae III
TGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTA  320
ACGTTGTATGCCTTTTGAATGGGAATTTAAATAAACGTGATGACCTTTTGATGGACAAGGTACCGGTTGTGAACAGTGAT
                     254                                      299                314
                                                              299
                                                              299
                                                               300
                                                                 302
                                                                 302
                                                                 302
                                                                  303

                                         Nde I
                                        Sau3A I
                                        Mbo I
Ser65→Thr65                             Dpn I           Nla III             Nla III
CTTTCaCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACgGCATGACTTTTTCAAGAGTGCCATGCCC  400
GAAAGTGAATACCACAAGTTACGAAAAGTTCTATGGGTCTAGTATACTTTGcCGTACTGAAAAAGTTCTCACGGTACGGG
                                        359             374                 394
                                        359
                                        359
                                           362

                                                           Mae II
         Rsa I                                           Afl III                Hph I
GAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGG  480
CTTCCAATACATGTCCTTTCTTGATATAAAAAGTTTCTACTGCCCTTGATGTTCTGTGCACGACTTCAGTTCAAACTTCC
         410                                             455                   479
```

Fig. 5C

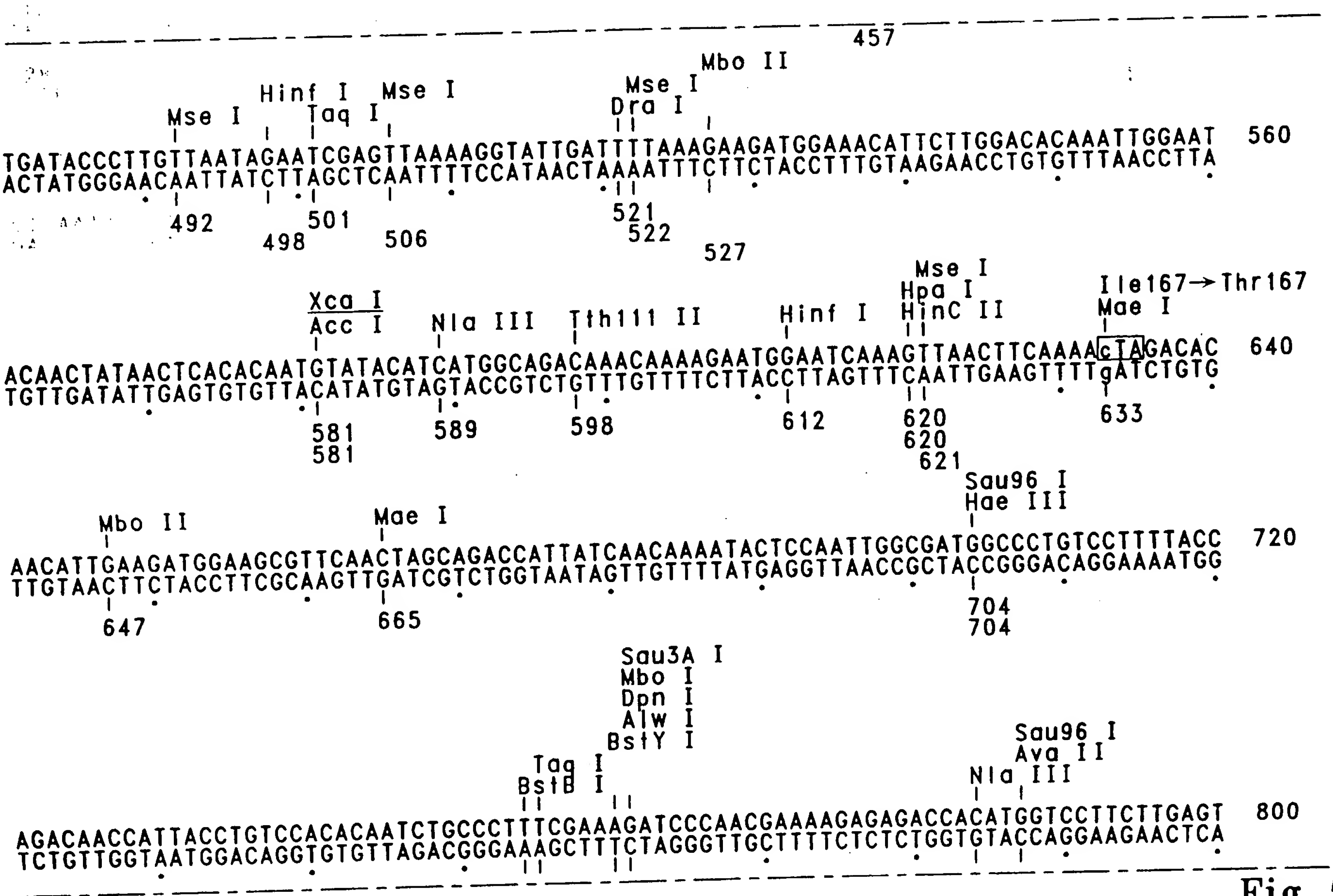

457
Mbo II
Mse I
Dra I
Hinf I
Taq I
Mse I
Mse I
TGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAAT 560
ACTATGGGAACAATTATCTTAGCTCAATTTTCCATAACTAAAATTTCTTCTACCTTTGTAAGAACCTGTGTTTAACCTTA
492
498
501
506
521
522
527
Mse I
Hpa I
HinC II
Ile167→Thr167
Mae I
Xca I
Acc I
Nla III
Tth111 II
Hinf I
ACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAcTAGACAC 640
TGTTGATATTGAGTGTGTTACATATGTAGTACCGTCTGTTTGTTTTCTTACCTTAGTTTCAATTGAAGTTTTgATCTGTG
581
581
589
598
612
620
620
621
633
Sau96 I
Hae III
Mbo II
Mae I
AACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACC 720
TTGTAACTTCTACCTTCGCAAGTTGATCGTCTGGTAATAGTTGTTTTATGAGGTTAACCGCTACCGGGACAGGAAAATGG
647
665
704
704
Sau3A I
Mbo I
Dpn I
Alw I
BstY I
Taq I
BstB I
Sau96 I
Ava II
Nla III
AGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGT 800
TCTGTTGGTAATGGACAGGTGTGTTAGACGGGAAAGCTTTCTAGGGTTGCTTTTCTCTCTGGTGTACCAGGAAGAACTCA

Fig. 5D

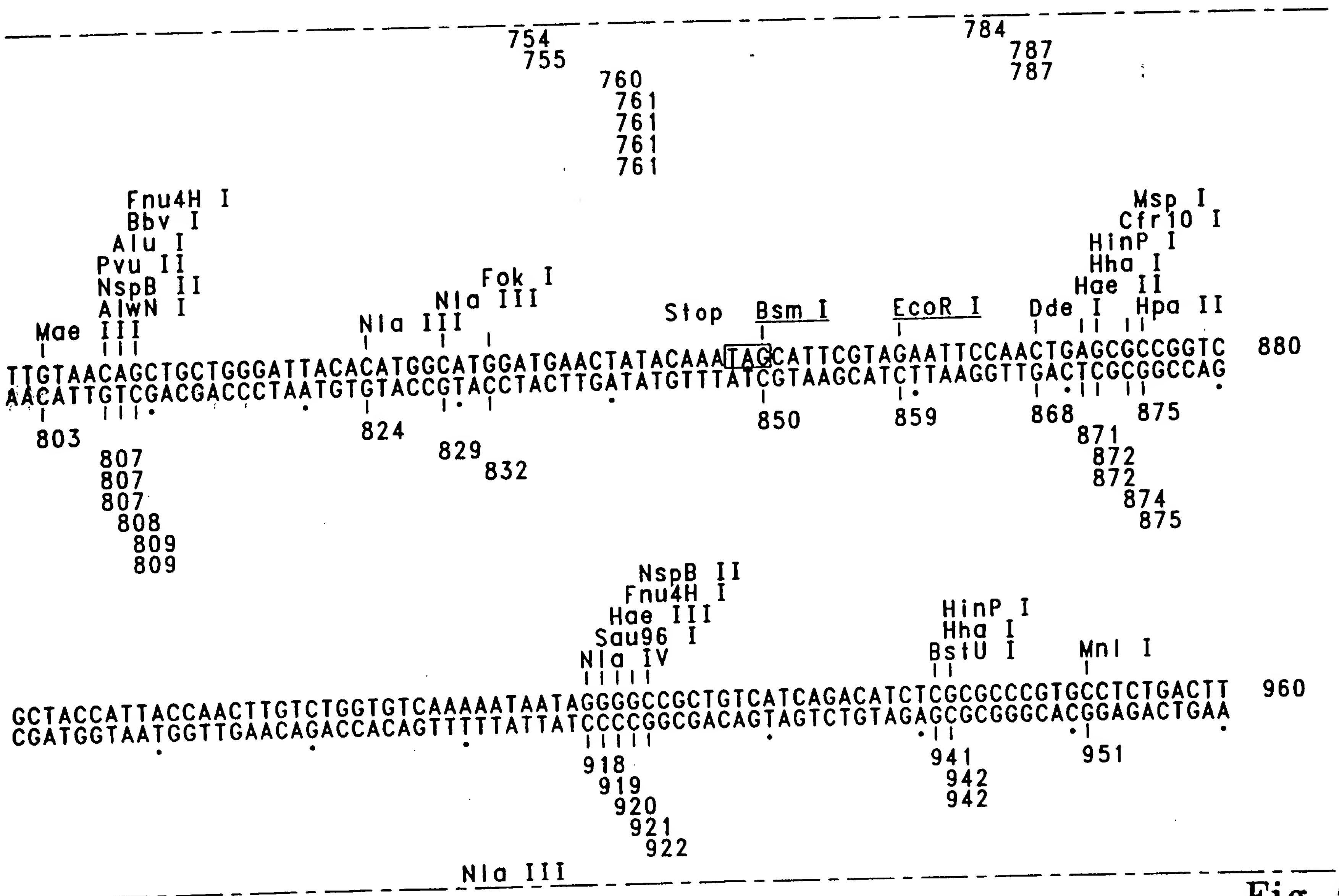

754
755
760
761
761
761
761
784
787
787
Fnu4H I
Bbv I
Alu I
Pvu II
NspB II
AlwN I
Mae III
Nla III
Nla III
Fok I
Stop
Bsm I
EcoR I
Dde I
Hae II
Hha I
HinP I
Cfr10 I
Msp I
Hpa II
TTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAGCATTCGTAGAATTCCAACTGAGCGCCGGTC
AACATTGTCGACGACCCTAATGTGTACCGTACCTACTTGATATGTTTATCGTAAGCATCTTAAGGTTGACTCGCGGCCAG
880
803
807
807
807
808
809
809
824
829
832
850
859
868
871
872
872
874
875
875
NspB II
Fnu4H I
Hae III
Sau96 I
Nla IV
HinP I
Hha I
BstU I
Mnl I
GCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGACATCTCGCGCCCGTGCCTCTGACTT
CGATGGTAATGGTTGAACAGACCACAGTTTTTATTATCCCCGGCGACAGTAGTCTGTAGAGCGCGGGCACGGAGACTGAA
960
918
919
920
921
922
941
942
942
951
Nla III

Fig. 5E

```
Dde I           Mbo II    Fok I   NspH I                HgiA I
                                  Nsp7524 I             Bsp1286 I
CTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAA  1040
GATTCAGGTTAATGAGAAGTTGTAGGGATGTACGAGAAAGAGGGACACGAGGGTGGGGGATAAAAACAATAATAGTTTTT
961             975       982     989                   1006
                                  989                   1006
                                  990

   Mbo II  Mse I        Alu I   Mse I Hph I  Mnl I                 Hinf I     Ase I
                                      Mae III                                 Xma I
AACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAA  1120
TTGAAGAAGAATTAAAGAAACAAAAAATCGAAGAAAATTCAGTGGAGATTGTTACTTTAACACATCTAAGTTTTTATCTT
   1045    1050         1068    1076  1080   1084                  1106       1118
                                      1081                                    1120

Mse I           Taq I        HgiA I   Mnl I     Mse I
                             Bsp1286 I          Ase I
TTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATG  1200
AATTAAGCATTATTTTTCAGCTTTTTTTAACACGAGGGAGGGGGGTAATTATTATTAAGATAGGGTTTTAGATGTGTTAC
1121            1139         1151     1157      1166
                             1151               1167

      Rsa I
TTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATA  1280
AAGACACATGTGAAGAATACAAAAAAAATGAAGACTATTTAAAAAAAACTTTGTAGTATCTTTTTTGGCGTGTGTTTTAT
      1207

Nde I  Mae III  Mae II    Mbo II  Mae II  Hae III  Mnl I  Sau96 I  Nla III  BspH I
Mae II  Aha II  Aat II  Nla III
```

Fig. 5F

```
CCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCA 1360
GGAATAGTATACAATGCAAAGTCAAATACTGGCGTTAAAAATAAAGAAGCGTGCAGACCCGGAGAGTACTGCAGTTTAGT
1287 1292 1295 1325 1332 1338 1339 1341 1349 1349 1350 1359
1345 1346
Mse I
Ase I

TGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTT 1440
ACGAGTAGCACTTTTTCAAAACCTCATAAAAACCTTAAAAAGTTAGTTCACTTTCAAATACTTTAATTAAAAGGACGAAA
1424
1425

Tth111 II   Mnl I   Taq I   HgiA I   Bsp1286 I
TGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATG 1520
ACGAAAAACCCCCAAAGGGGATAACAAACAGTTCTCAAAGCTCCTGCCGCAAAAAGAACGATTTTAGTGTTCATAACTAC
1464   1479 1481   1520 1520

SfaN I   Mbo II   Sau3A I   Mbo I   Dpn I   Mnl I   Dde I   Hph I
AGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGT 1600
TCGTGCTACGTTCTTTCTAGCCTTCTTCCAAACCCAAACTCCGAGTCACCTTCCACTCATCTTCAACTATTAAACTTTCA
1526   1537 1537 1537 1542   1559 1563   1573

Gdi II
Eag I
Eae I
```

Fig. 5G

Mse I Xmn I Tth111 II Mae I Spe I

```
GGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTC  1680
CCTCATCACAGATACCCCAAAAACGGAATTTACTGTCTTATGTAAGGGTTATATGGTTTGTATTGACAAAGGATGATCAG
```

1627 1637 1656 1674 1675 1680 1680 1680

Hae III Sfi I Bgl I Spl I Rsa I Hae III Apa I Ban II Bsp1286 I Eco0109 I Nla IV Sau96 I Sau96 I BstU I Hha I HinP I BstU I Hph I Hph I Mnl I Nsp7524 I NspH I Nla III Bbv I Fnu4H I Alu I ScrF I Hpa II Bcn I Nci I Msp I Mae III

```
GGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTC  1760
CCGGCATGCCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAG
```

1681 1681 1682 1684 1685 1690 1689 1689 1689 1689 1689 1689 1690 1703 1704 1704 1705 1713 1722 1730 1738 1738 1739 1742 1742 1744 1749 1748 1748 1749 1748 1758

ScrF I Nci I Msp I Hpa II Bcn I BstU I Hga I NspB II

Fig. 5H

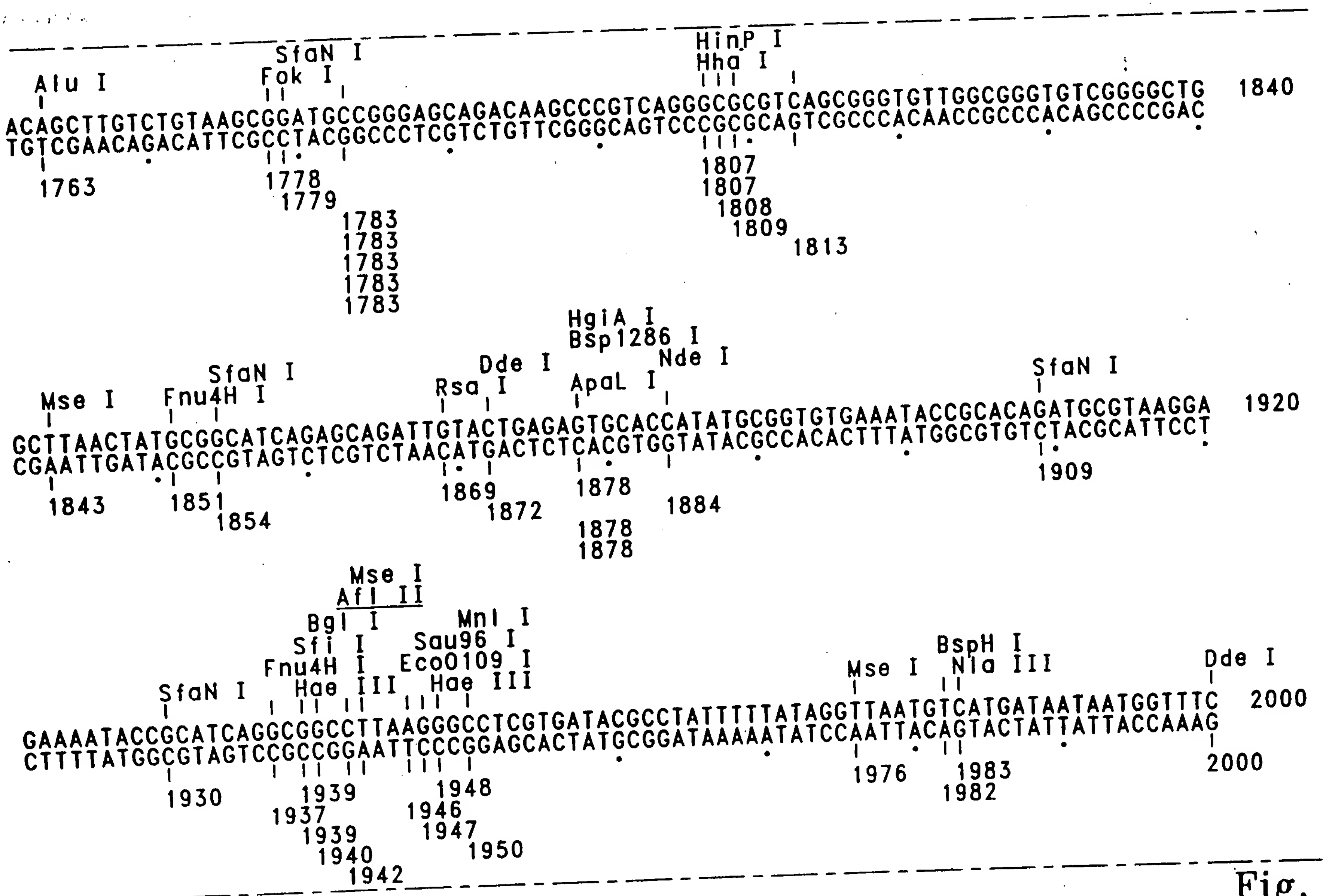

SfaN I
Alu I
Fok I
HinP I
Hha I
ACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTG
TGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGAC
1840
1763
1778
1779
1783
1783
1783
1783
1783
1807
1807
1808
1809
1813
HgiA I
Bsp1286 I
Nde I
Dde I
SfaN I
SfaN I
Mse I
Fnu4H I
Rsa I
ApaL I
GCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGA
CGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCCT
1920
1843
1851
1854
1869
1872
1878
1878
1878
1884
1909
Mse I
Afl II
Bgl I
Mnl I
Sfi I
Sau96 I
Fnu4H I
Eco0109 I
BspH I
Nla III
Dde I
SfaN I
Hae III
Hae III
Mse I
GAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTC
CTTTTATGGCGTAGTCCGCCGGAATTCCCGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTACCAAAG
2000
2000
1930
1937
1939
1939
1940
1942
1946
1947
1948
1950
1976
1982
1983

Fig. 5I

```
                                   Nla IV
                                 BstU I
     Mae II                     HinP I
     Aha II                     Hha I
     Aat II                     ||    |
     ||
TTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT  2080
AATCTGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACA
     ||                          ||    |
     2004                        2034
     2004                        2034
      2005                        2035
                                    2038

                                                  Mbo II
    Nla III                   Ssp I               |
   BspH I                     |
   ||          |
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGT  2160
TAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCA
   ||          |              |                   |
   2087                       2120                2133
    2088

                                                               Hph I              SfaN I
                                       Hph I                   |                  |
               Fnu4H I                 |
               |
GTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGC  2240
CAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCTACG
               |                       |                       |                  |
               2182                    2208                    2223               2236

                                                          Sau3A I
                                     Sau3A I              Mbo I
                                     Mbo I                Dpn I
  Sau3A I                            Dpn I                BstY I
  Mbo I        HgiA I                BstY I               Alw I                   Mbo II
  Dpn I        Bsp1286 I       Taq I Alw I      NspB II   ||                      |
Mbo II         ApaL I    Mae III     ||         |
|   |          |         |     |
TGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCG  2320
ACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGGC
|   |          |         |     |     ||         |         ||                      |
2242           2255      2267  2273  2280       2289      2298                    2320
    2245       2255                  2280                 2297
    2245       2255                  2281                 2298
    2245                             2281                 2298
```

Fig. 5J

```
2281  2298

Xmn I  Mae II  HgiA I  Bsp1286 I  Dra I  Mse I  BstU I  HinP I  Hha I  Hga I  Aha II  ScrF I  Nci I  Msp I  Hpa II  Bcn I
AAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAG 2400
TTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTC
2323 2325  2340 2340  2348 2347  2366 2366 2367  2386 2386  2390 2390 2390 2390 2390

Fnu4H I  Dde I  Rsa I  Sca I  Hph I  Mae III  SfaN I  Fok I
CAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGG 2480
GTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACC
2411  2426  2444 2445 2449  2455  2466  2475

Nla III  Fnu4H I  Bbv I  Nla III  Fnu4H I  Eae I  Gdi II  Hae III  Pvu I  Dpn I  Mbo I  Sau3A I
CATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG 2560
GTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGC
2481  2506 2506  2517  2533 2534 2534  2555 2556 2556
```

Fig. 5K

```
                                                         2535                          2556
                                   Sau3A I
                                   Mbo I
                                   Dpn I                    Sau3A I        Msp I
   Sau96 I                        Alw I  Mae III            Mbo I          Hpa II
   Ava II                   Nla III  Nla III                Dpn I   Nla IV  Alu I
Mnl I       Alu I
GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTG  2640
CTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACCCTTGGCCTCGAC
2561        2573            2595   2601 2605 2608            2620   2628  2632 2636
 2563                                2602                       2620         2632
 2563                                2602                       2620
                                     2602

                                                          HinP I
                                                          Hha I
                                                         Fsp I
                        Mae III  SfaN I              Mae  II           Mse I
AATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGG  2720
TTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGATAATTGACC
                        2666     2676                2698              2713
                                                        2702
                                                         2703
                                                         2703

               Msp I
               Hpa II
              ScrF I
              Nci I
          Alu I          Mse I                                                HinP I
        Mae I  Bcn I     Ase I     Fok I  Mnl I               Sau96 I         Hha I
                                                              Ava II
CGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCT  2800
GCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGA
        2734   2741      2751      2762   2767                2785           2796
         2736             2752                                2785           2796
               2741
```

Fig. 5L

```
                         2741
                          2742
                          2742
           Msp I
           Hpa II                                        Hph I
  Bgl I                                                Msp I
 Sau96 I                                               Hpa II                              Fnu4H I
 Hae III                                              Cfr10 I        BstU I              Bbv I      Nla IV
                                                   Nla IV
CGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTG  2880
GCCGGGAAGGCCGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGAC
 2802                                      2839              2860        2872      2880
 2802                                        2842                        2872
  2803                                         2843
                                               2843
           2809                                  2845
           2809                                    Ple I
 Mae III                                           Hinf I            Fok I
Sau96 I            Mnl I
GGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACA  2960
CCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGT
2881               2897                            2928              2943
 2882                                              2928
    Dde I
Sau3A I                Mnl I                                                      Mse I
Mbo I              Nla IV                                                        Dra I
Dpn I              Ban I        Mse I      Mae III
GATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATT  3040
CTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTAA
2961               2974         2987       2997                                   3039
2961               2974                                                            3040
2961                   2978
     2966
                                         Sau3A I
                                Mae I    Mbo I
                              Sau3A I    Dpn I
```

Fig. 5M

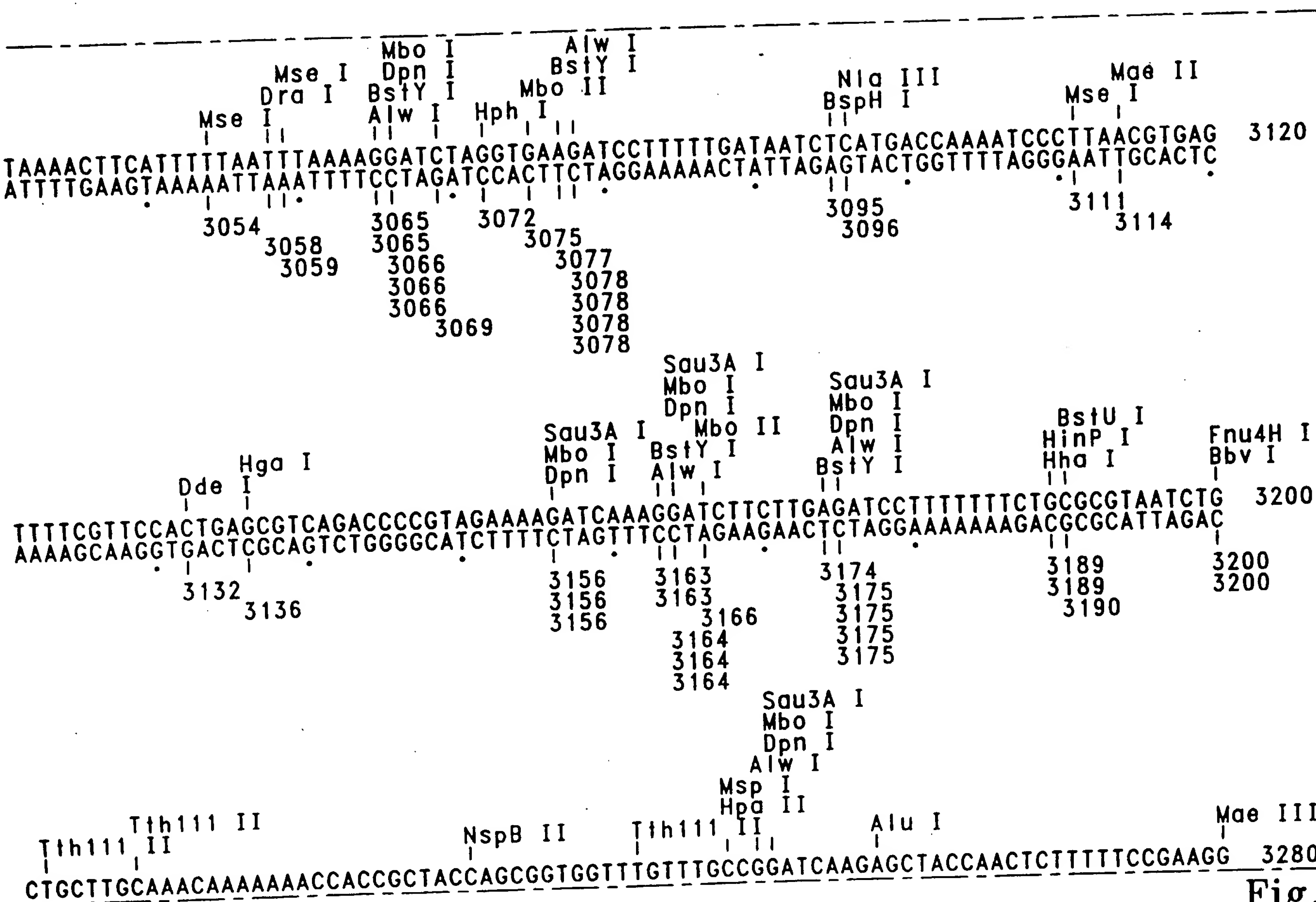

Mse I
Dra I
Mse I
Mbo I
Dpn I
BstY I
Alw I
Hph I
Mbo II
Alw I
BstY I
Nla III
BspH I
Mse I
Mae II
TAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG 3120
ATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTC
3054
3058
3059
3065
3065
3066
3066
3066
3069
3072
3075
3077
3078
3078
3078
3078
3095
3096
3111
3114
Dde I
Hga I
Sau3A I
Mbo I
Dpn I
Sau3A I
Mbo I
Dpn I
Mbo II
BstY I
Alw I
Sau3A I
Mbo I
Dpn I
Alw I
BstY I
BstU I
HinP I
Hha I
Fnu4H I
Bbv I
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTG 3200
AAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGAC
3132
3136
3156
3156
3156
3163
3163
3166
3164
3164
3164
3174
3175
3175
3175
3175
3189
3189
3190
3200
3200
Tth111 II
Tth111 II
NspB II
Tth111 II
Sau3A I
Mbo I
Dpn I
Alw I
Msp I
Hpa II
Alu I
Mae III
CTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG 3280

Fig. 5N

```
GACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCC
3202  3208          3230     3241  3247 3247 3249 3250 3250 3250  3257        3280

HinP I, Hha I (3298, 3298); Mae I (3322); Hae III, Hae I (3339, 3340)
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT   3360
ATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGA

Mnl I (3378); Mae III (3396); AlwN I (3401); Fnu4H I, Bbv I (3406, 3406); Fnu4H I, Bbv I (3409, 3409); ScrF I, Nci I, Msp I, Hpa II, Bcn I (3437, 3437, 3437, 3437, 3437)
GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG   3440
CATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCC

Ple I, Hinf I; Mae III; Msp I, Hpa II; HinP I, Hha I, Bbv I, Fnu4H I, NspB II; HgiA I, Bsp1286 I, ApaL I; Alu I
GTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG   3520
CAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACC
```

Fig. 50

```
            Hae III            EcoR II                    NspH I
          Fnu4H I              BstN I          Hae III    Nsp7524 I
         BstU I        Nla     IV             Hae I       Afl III
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCT  3840
TGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGA
         3771          3785            3802       3815
          3772           3789           3803      3815
            3774         3789                     3815
                         3789                      3816
                           3791
                            3792

                                                                                 Ple I
                                                                            Fnu4H I
                                                          Fnu4H I           Bbv I
                                                          Bbv I           HinP I
Hinf I                                  Alu I           Fnu4H I           Hha I  Hinf I
GATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTC  3920
CTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCAG
3841                                    3876            3890              3909   3916
                                                          3893            3909
                                                          3893              3911
                                                                            3911
                                                                                 3916

                                                                 Gdi II
                                                                 Eae I
                                                            BstU I             Pvu II
                        HinP I                             HinP I            Fnu4H I
                        Hha I                              Hha I      Mse I Alu I
                       Hae II                              Hha I    Hinf I   Bbv I
       Mnl I     Mbo II                     Mnl I         BstU I  Hae III  Ase I  NspB II
AGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC  4000
TCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCG
       3928      3937                       3961          3969   3977     3986  3993
                         3941                              3970      3981      3992
                          3942                             3970         3987    3994
```

Fig. 5Q

```
                                3942                                          3971            3992
                                                                                 3976          3993
                                                                                 3976
                                                                                                 ScrF I
                                                                                                 EcoR II
                                                                                                 BstN I
                                                                                                Sec I
                                              HinP I       Mse I                             Nla IV
                                              Hha I       Ase I            Alu I             Ban I
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG  4080
TGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTC
                                              4035         4045            4058             4071
                                              4035          4046                            4071
                                                                                                4076
                                                                                                 4077
                                                                                                 4077
                                                                                                 4077

                         Msp I                                                            Alu I
                         Hpa II
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGA  4160
CGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACT
                         4099                                                             4153
                         4099

 Nla III
CCATGATTACGCC  4173
GGTACTAATGCGG
 4162
```

Fig. 5R

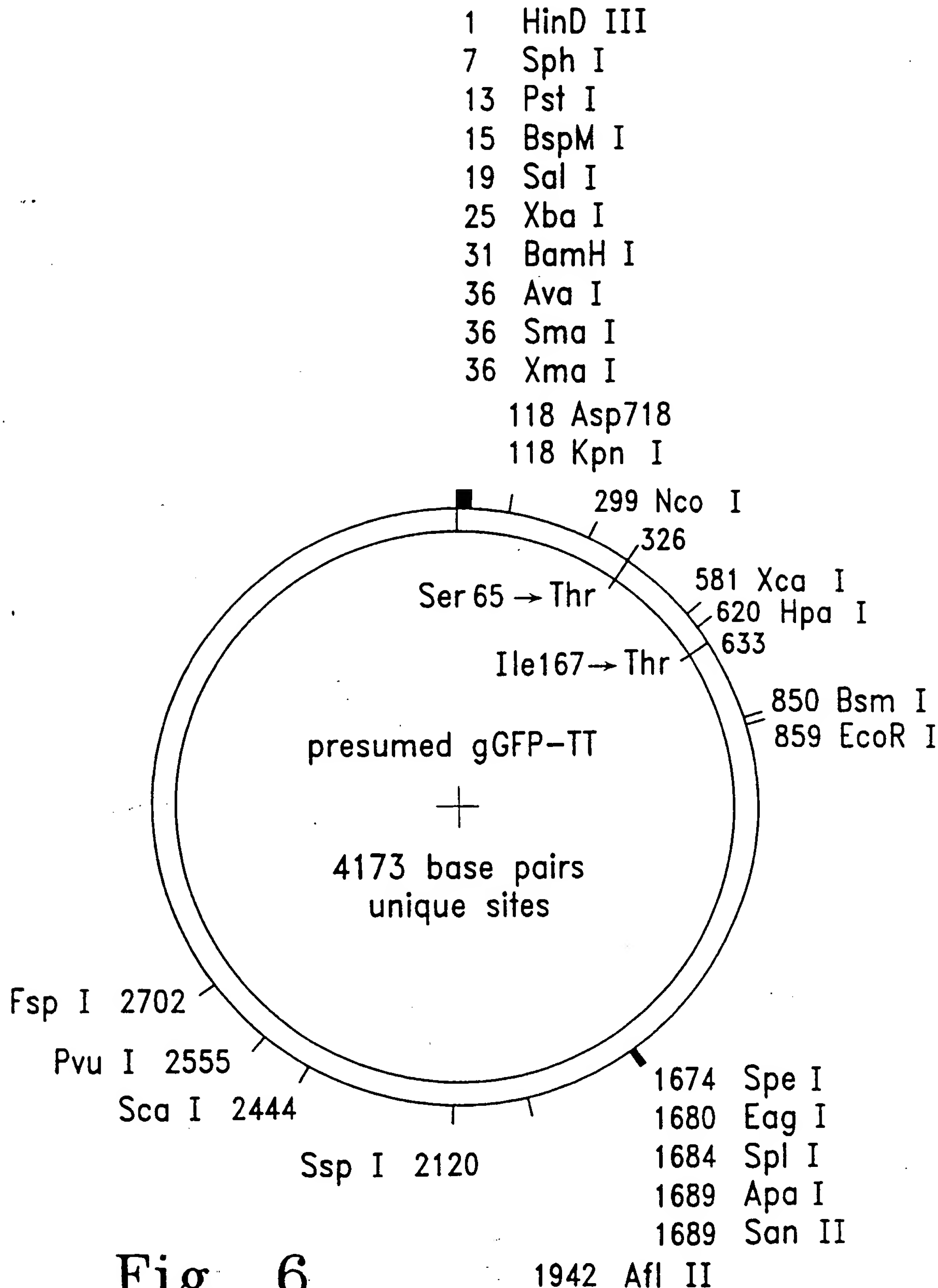

1 HinD III
7 Sph I
13 Pst I
15 BspM I
19 Sal I
25 Xba I
31 BamH I
36 Ava I
36 Sma I
36 Xma I
118 Asp718
118 Kpn I
299 Nco I
326
Ser 65 → Thr
581 Xca I
620 Hpa I
633
Ile167 → Thr
850 Bsm I
859 EcoR I
presumed gGFP-TT
4173 base pairs
unique sites
Fsp I 2702
Pvu I 2555
Sca I 2444
Ssp I 2120
1674 Spe I
1680 Eag I
1684 Spl I
1689 Apa I
1689 San II
1942 Afl II

Fig. 6

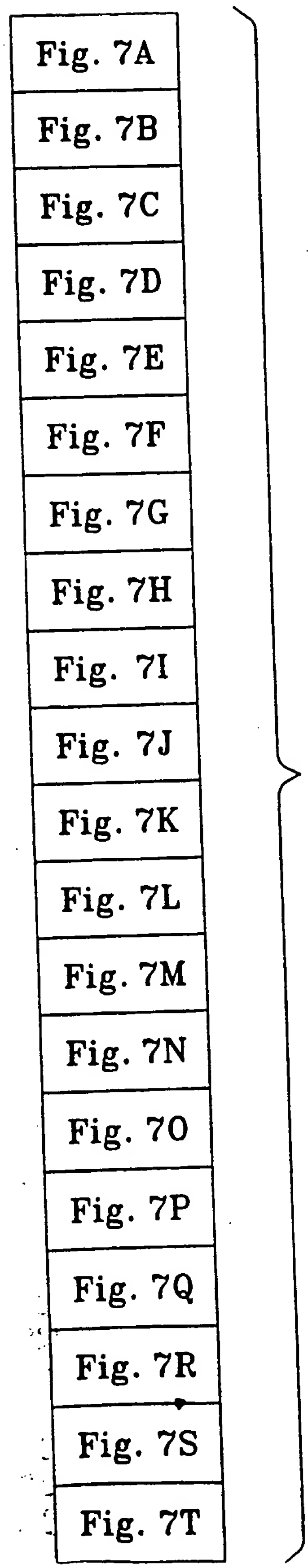


Fig. 7A
Fig. 7B
Fig. 7C
Fig. 7D
Fig. 7E
Fig. 7F
Fig. 7G
Fig. 7H
Fig. 7I
Fig. 7J
Fig. 7K
Fig. 7L
Fig. 7M
Fig. 7N
Fig. 7O
Fig. 7P
Fig. 7Q
Fig. 7R
Fig. 7S
Fig. 7T

Fig. 7

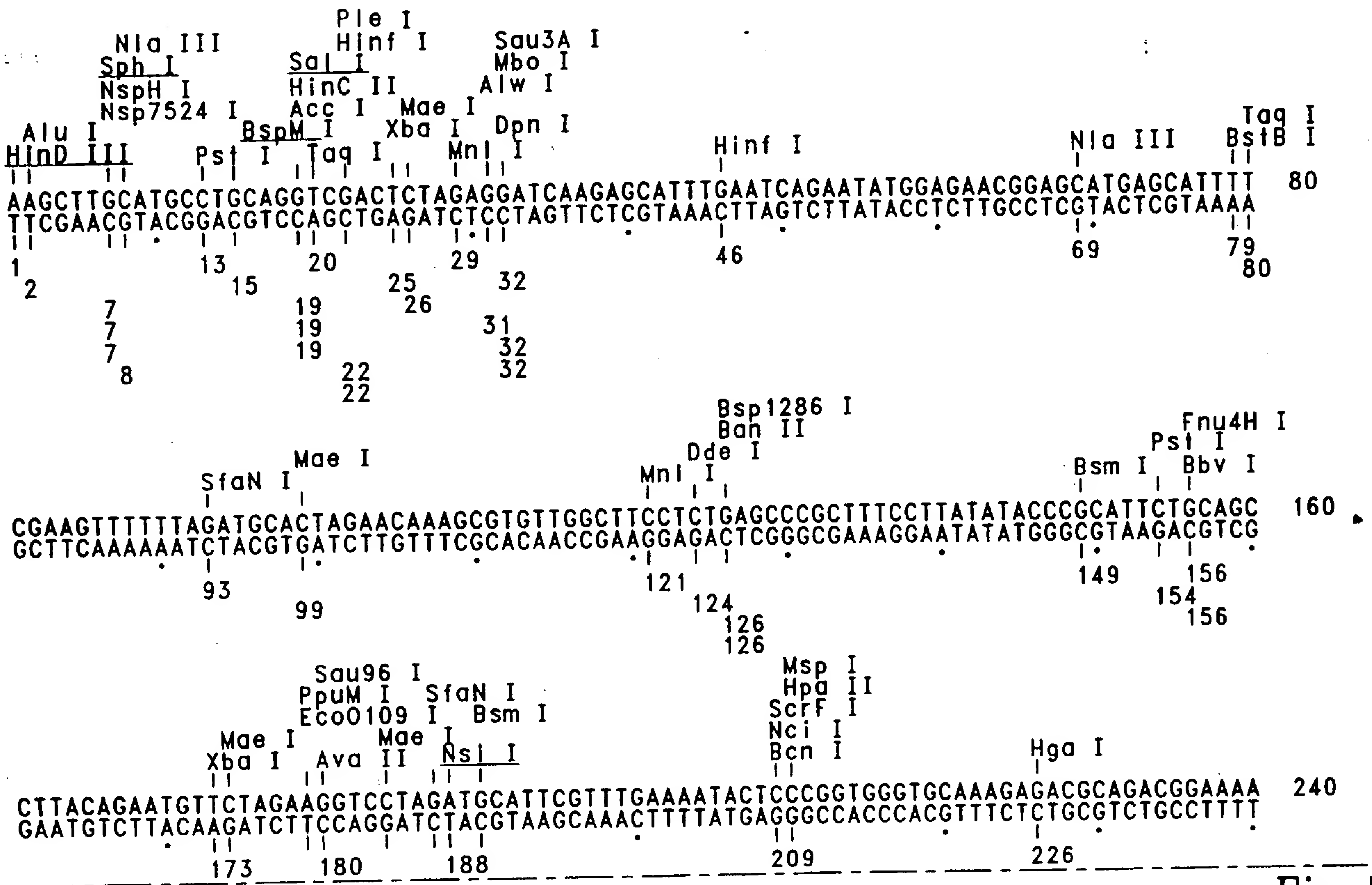

Ple I
Hinf I
Nla III
Sph I
NspH I
Nsp7524 I
Alu I
HinD III
Pst I
BspM I
Sal I
HinC II
Acc I
Taq I
Mae I
Xba I
Sau3A I
Mbo I
Alw I
Dpn I
Mnl I
Hinf I
Nla III
Taq I
BstB I
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCAAGAGCATTTGAATCAGAATATGGAGAACGGAGCATGAGCATTTT 80
TTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGTTCTCGTAAACTTAGTCTTATACCTCTTGCCTCGTACTCGTAAAA
1 2 7 7 7 8 13 15 20 19 19 19 22 22 25 26 29 32 31 32 32 46 69 79 80
Bsp1286 I
Ban II
Dde I
Mnl I
SfaN I
Mae I
Bsm I
Pst I
Fnu4H I
Bbv I
CGAAGTTTTTTAGATGCACTAGAACAAAGCGTGTTGGCTTCCTCTGAGCCCGCTTTCCTTATATACCCGCATTCTGCAGC 160
GCTTCAAAAAATCTACGTGATCTTGTTTCGCACAACCGAAGGAGACTCGGGCGAAAGGAATATATGGGCGTAAGACGTCG
93 99 121 124 126 126 149 154 156 156
Sau96 I
PpuM I
Eco0109 I
SfaN I
Bsm I
Mae I
Xba I
Ava I
Mae II
Mae I
Nsi I
Msp I
Hpa II
ScrF I
Nci I
Bcn I
Hga I
CTTACAGAATGTTCTAGAAGGTCCTAGATGCATTCGTTTGAAAATACTCCCGGTGGGTGCAAAGAGACGCAGACGGAAAA 240
GAATGTCTTACAAGATCTTCCAGGATCTACGTAAGCAAACTTTTATGAGGGCCACCCACGTTTCTCTGCGTCTGCCTTTT
173 180 188 209 226

Fig. 7A

```
            174       184          209
               179         190     209
               179      187         210
                180                 210
                                                                        Taq I
                                                                     Alu I
                                                                  Fnu4H I
                                                                  Bbv I
                              Rsa I      Ple I  Mnl I             Tth111 II
                   Rsa I   Nla III       Hinf I
TGTATCTGGGTCTCTTTATTGTGTACACTACTTTTCCATGTACCGAATGTGAGTCGCCCTCCTTTTGCAACAAGCAGCTC  320
ACATAGACCCAGAGAAATAACACATGTGATGAAAAGGTACATGGCTTACACTCAGCGGGAGGAAAACGTTGTTCGTCGAG
                     263        277 280     291   298             311
                                            291                     314
                                                                    314
                                                                      316
                                                                        319

     Mae I
    Xba I
GAATGTTCTAGAAAAAGGTGGAAAATAGTATAAATACCGTTGAAAATAAATACCGAACAACATTTGCTCTAATTGTGAAA  400
CTTACAAGATCTTTTTCCACCTTTTATCATATTTATGGCAACTTTTATTTATGGCTTGTTGTAAACGAGATTAACACTTT
     327
      328
                              ScrF I
                              Nci I
                              Msp I
                              Hpa II
                              Bcn I
                             Xma I
                             Sma I
                             Sec I
                             ScrF I
                             Nci I
                             Bcn I
                             Ava I
                            Sec I
                          Sau3A I
```

Fig. 7B

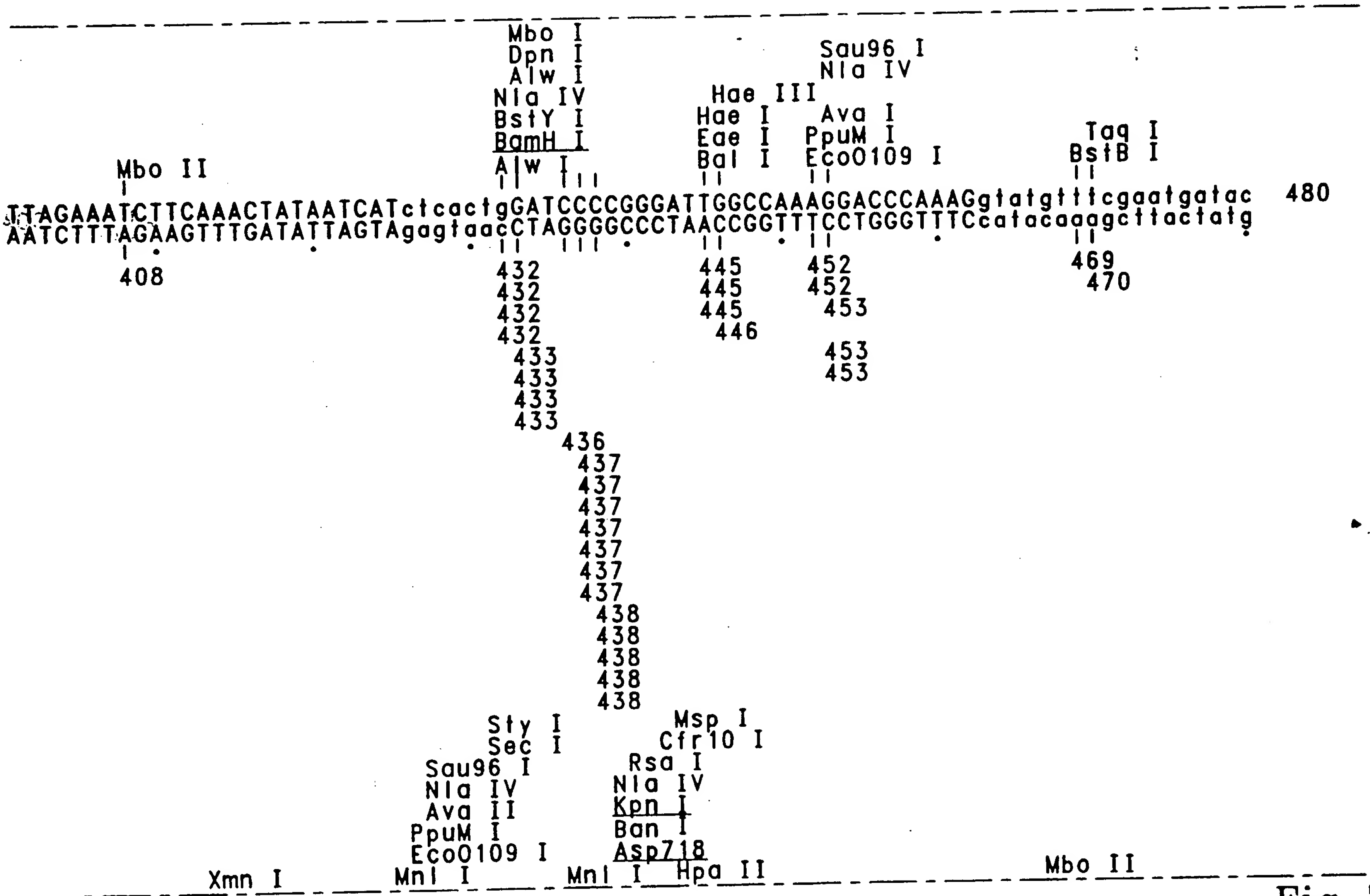

Mbo II
Mbo I
Dpn I
Alw I
Nla IV
BstY I
BamH I
Alw I
Hae III
Hae I
Eae I
Bal I
Sau96 I
Nla IV
Ava I
PpuM I
EcoO109 I
Taq I
BstB I
TTAGAAATCTTCAAACTATAATCATctcactgGATCCCCGGGATTGGCCAAAGGACCCAAAGgtatgtttcgaatgatac 480
AATCTTTAGAAGTTTGATATTAGTAgagtaacCTAGGGGCCCTAACCGGTTTCCTGGGTTTCcatacaaagcttactatg
408
432
432
432
432
433
433
433
433
436
437
437
437
437
437
437
437
438
438
438
438
438
445
445
445
446
452
452
453
453
453
469
470
Sty I
Sec I
Sau96 I
Nla IV
Ava II
PpuM I
EcoO109 I
Msp I
Cfr10 I
Rsa I
Nla IV
Kpn I
Ban I
Asp718
Xmn I
Mnl I
Mnl I
Hpa II
Mbo II

Fig. 7C

```
                                                                                    Mcl
taacataacatagaacattttcagGAGGACCCTTGGAGGGTACCGGTAGAAAAAATGAGTAAAGGAGAAGAACTTTTCAC   560
attgtattgtatcttgtaaaagtcCTCCTGGGAACCTCCCATGGCCATCTTTTTTACTCATTTCCTCTTCTTGAAAAGTG
           493           505      516  523                            547
                          506           519
                          506           519
                           507          519
                           507          519
                           507           520
                               511        522
                               511         523
                                            Bsp1286 I

                               Hph I   Mse I                                 Mnl I Hph I Hph I SfaN I
TGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTG   640
ACCTCAACAGGGTTAAGAACAACTTAATCTACCACTACAATTACCCGTGTTTAAAAGACAGTCACCTCTCCCACTTCCAC
                               592    599    604                          628  631  637  640

                                                                   Hae III
                                                                   Hae I
                                                                   Eae I
                                                                   Bal I
                                                                   Nla III
                                                                   Sty I
                                                                   Sec I
                         Mse I                                     Nco I              Mae III
ATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACT   720
TACGTTGTATGCCTTTTGAATGGGAATTTAAATAAACGTGATGACCTTTTGATGGACAAGGTACCGGTTGTGAACAGTGA
                         665                                       700                 715
                                                                   700
                                                                   700
                                                                    701
                                                                     703
                                                                     703
                                                                     703
                                                                      704
```

Fig. 7D

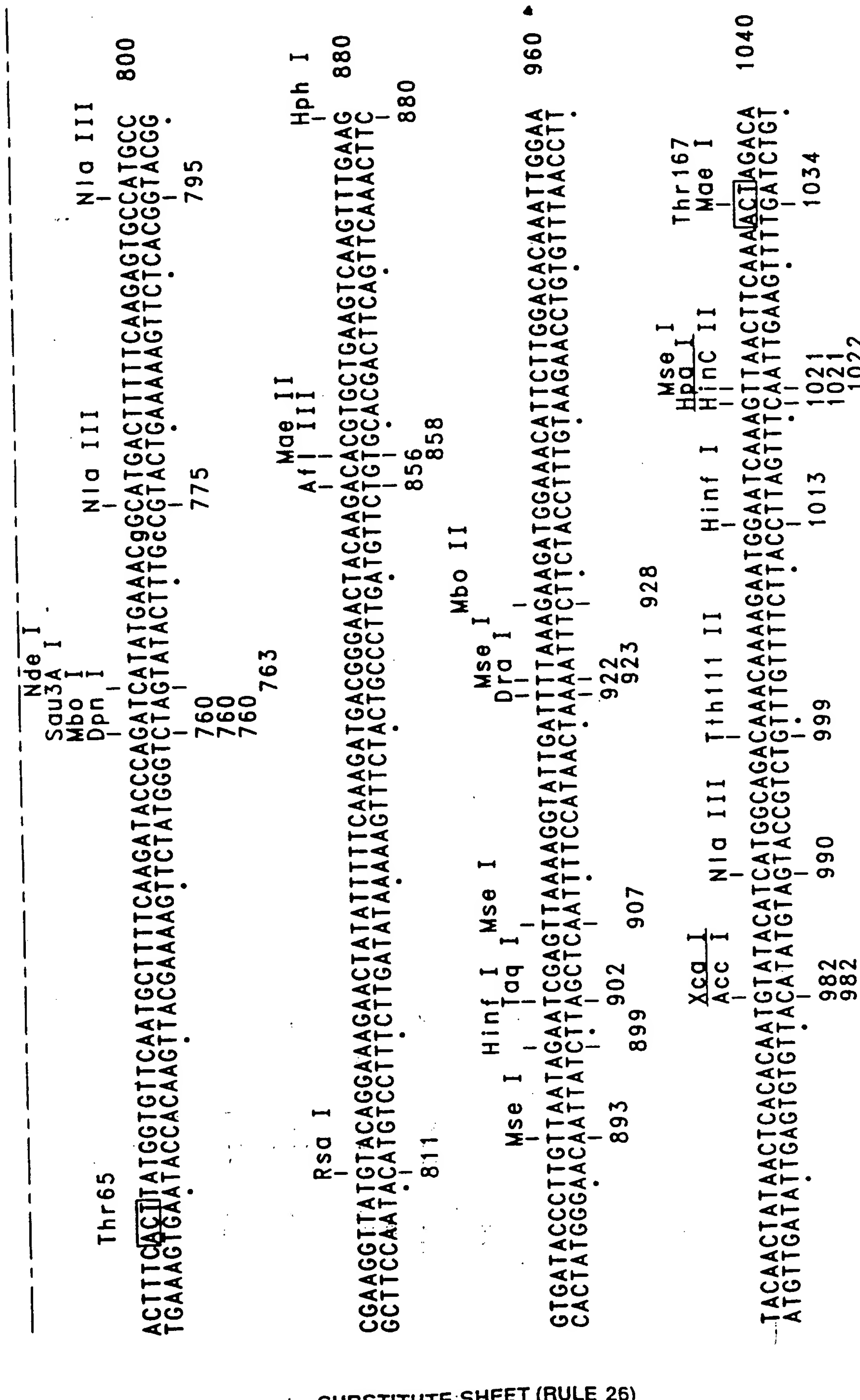


Fig. 7E

```
Mbo II                           Mae I                                                                 Sau96 I
                                                                                                       Hae III
CAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTAC   1120
GTTGTAACTTCTACCTTCGCAAGTTGATCGTCTGGTAATAGTTGTTTTATGAGGTTAACCGCTACCGGGACAGGAAAATG
1048                             1066                                                                  1105
                                                                                                       1105

                                                  Sau3A I
                                                  Mbo I
                                                  Dpn I
                                                  Alw I
                                                  BstY I                       Sau96 I
                                         Taq I                                 Ava II
                                         BstB I                              Nla III
CAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAG   1200
GTCTGTTGGTAATGGACAGGTGTGTTAGACGGGAAAGCTTTCTAGGGTTGCTTTTCTCTCTGGTGTACCAGGAAGAACTC
                                         1155     1161                       1185
                                          1156     1162                         1188
                                                   1162                         1188
                                                   1162
                                                   1162

             Fnu4H I                                                                                 Msp I
             Bbv I                                                                                   Cfr10 I
            Alu I                                                                                   HinP I
            Pvu II                                                                                  Hha I
            NspB II                                 Fok I                                          Hae II
            AlwN I                        Nla III          Stop                                     
     Mae    III                    Nla III                 Bsm I       EcoR I          Dde I      Hpa II
TTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAGCATTCGTAGAATTCCAACTGAGCGCCGGT   1280
AAACATTGTCGACGACCCTAATGTGTACCGTACCTACTTGATATGTTTATCGTAAGCATCTTAAGGTTGACTCGCGGCCA
     1204                          1225  1230      1251     1260       1269     1276
            1208                             1233                             1272
            1208                                                               1273
```

Fig. 7F

52/68

```
     1208                                                                        1273
      1209                                                                         1275
       1210                                                                         1276
       1210
                                        NspB II
                                       Fnu4H I
                                      Hae III                    HinP I
                                     Sau96 I                     Hha I
                                    Nla IV                      BstU I    Mnl I
                                    |||||                        ||        |
CGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGACATCTCGCGCCCGTGCCTCTGACT  1360
GCGATGGTAATGGTTGAACAGACCACAGTTTTTATTATCCCCGGCGACAGTAGTCTGTAGAGCGCGGGCACGGAGACTGA
                                    |||||                        ||        |
                                    1319                         1342      1352
                                     1320                         1343
                                      1321                        1343
                                       1322
                                        1323
                              Nla III
                    Fok I     NspH I                Hgi A I
 Dde I       Mbo II           Nsp7524 I             Bsp1286 I
 |           |      |         ||                    |
TCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAA  1440
AGATTCAGGTTAATGAGAAGTTGTAGGGATGTACGAGAAAGAGGGACACGAGGGTGGGGGATAAAAACAATAATAGTTTT
 |           |      |         ||                    |
 1362        1376             1390                  1407
                    1383      1390                  1407
                               1391
                                            Hph I
               Mse I                       Mae III
     Mbo II                Alu I     Mse I       Mnl I                Hinf I     Xmn I
     |    |                |         |      ||   |                    |          |
AAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGA  1520
TTTGAAGAAGAATTAAAGAAACAAAAAATCGAAGAAAATTCAGTGGAGATTGTTACTTTAACACATCTAAGTTTTTATCT
     |    |                |         |      ||   |                    |          |
     1446                  1469      1477        1485                 1507       1519
          1451                              1481
                                             1482
                                        Mnl I
 Mse I                      HgiA I                  Mse I
```

Fig. 7G

```
Ase I              Taq I          Bsp1286 I       Ase I
ATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAAT  1600
TAATTAAGCATTATTTTTCAGCTTTTTTTAACACGAGGGAGGGGGGTAATTATTATTAAGATAGGGTTTTAGATGTGTTA
1521               1540           1552            1567
 1522                             1552             1568
                                       1558

      Rsa I
GTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAAT  1680
CAAGACACATGTGAAGAATACAAAAAAAATGAAGACTATTTAAAAAAAACTTTGTAGTATCTTTTTTGGCGTGTGTTTTA
      1608

                                                             Nla III
                                                            BspH I
                                                      Hae III      Mae II
                                              Mae II    Mnl I     Aha II
          Mae III                    Mbo II         Sau96 I       Aat II      Nla III
   Nde I   Mae II
ACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATC  1760
TGGAATAGTATACAATGCAAAGTCAAATACTGGCGTTAAAAATAAAGAAGCGTGCAGACCCGGAGAGTACTGCAGTTTAG
     1688   1696                     1726    1733   1739          1750        1760
         1693                                         1742        1750
                                                     1740          1751
                                                            1746
                                                             1747

                                                                Mse I
                                                               Ase I
ATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTT  1840
TACGAGTAGCACTTTTTCAAAACCTCATAAAAACCTTAAAAAGTTAGTTCACTTTCAAATACTTTAATTAAAAGGACGAA
                                                               1825
                                                                1826

                                   Mnl I
                     Tth111 II    Taq I
TTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGAT  1920
```

Fig. 7H

```
AACGAAAAACCCCCAAAGGGGATAACAAACAGTTCTCAAAGCTCCTGCCGCAAAAAGAACGATTTTAGTGTTCATAACTA
                  1865           1880
                                   1882

                  Mbo II
SfaN I     Sau3A I
HglA I     Mbo I                       Dde I
Bsp1286 I  Dpn I                   Mnl I         Hph I
GAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAG  2000
CTCGTGCTACGTTCTTTCTAGCCTTCTTCCAAACCCAAACTCCGAGTCACCTTCCACTCATCTTCAACTATTAAACTTTC
1921        1938                    1960       1974
1921        1938                       1964
   1927     1938
               1943

                                                                      Mae I
                  Mse I      Xmn I              Tth111 II             Spe I
TGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGT  2080
ACCTCATCACAGATACCCCAAAAACGGAATTTACTGTCTTATGTAAGGGTTATATGGTTTGTATTGACAAAGGATGATCA
                  2028       2038               2057                  2075
                                                                       2076

        Sau96 I
       Sau96 I
    Rsa I
   Spl I                                                                      ScrF I
  Bgl I    Nla IV                                                         Alu I
 Sfi I     EcoO109 I                                                  Fnu4H I
 Hae III    Hae III          BstU I                                  Bbv I   Msp I
Gdi II     Bsp1286 I        HinP I                                 Nla III   Ncl I
Eag I      Ban II           Hha I                                 NspH I     Bcn I
Eae I      Apa I           BstU I    Hph I     Hph I     Mnl I    Nsp7524 I  Hpa II   Mae III
CGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGT  2160
GCCGGCATGCCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCGTCTGCCA
2081       2090            2104      2114      2123      2131     2139       2150     2159
2081       2090             2105                                  2139       2149
```

Fig. 7I

```
2081 2082 2082 2083 2085 2086   2090 2091 2090 2090 2090 2091   2105 2106
2140 2143 2143 2145   2149 2150 2149

                                 ScrF I
                                 Nci I
                                 Msp I                                   NspB II
                                 Hpa II                          Hga I
                                 Bcn I                           BsfU I
                         SfaN I                                  HinP I
   Alu I                 Fok I                                   Hha I
CACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCT   2240
GTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGA
   2164                  2179 2180                               2208 2208 2209 2210 2214
                              2184 2184 2184 2184 2184

                                              HgiA I
                                              Bsp1286 I
                                       Dde I        Nde I
                SfaN I                                                  SfaN I
   Mse I    Fnu4H I                  Rsa I    ApaL I
GGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGG   2320
CCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCC
   2244     2252 2255                2270 2273     2279 2279 2279   2285            2310

                Mse I
                Afl II
```

Fig. 7J

```
SfaN I 2331; Fnu4H I 2338; Hae III 2340; Sfi I 2340; Bgl I 2341; Hae III 2343; 2344;
Eco0109 I 2347; Hae III 2348; Sau96 I 2349; Mnl I 2351; Mse I 2377; BspH I 2383; Nla III 2384

AGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTT   2400
TCTTTTATGGCGTAGTCCGCCGGAATTCCCGGAGCACTATGCGGATAAAAATATCCAATTACTGTACTATTATTACCAAA

Dde I 2401; Aat II 2405; Aha II 2405; Mae II 2406; Hha I 2435; HinP I 2435; BstU I 2436; Nla IV 2439

CTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATG   2480
GAATCTGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATAC

BspH I 2488; Nla III 2489; Ssp I 2521; Mbo II 2534

TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG   2560
ATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGC

Fnu4H I 2583; Hph I 2609; Hph I 2624; SfaN I 2637

TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATG   2640
ACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCTAC
```

Fig. 7K

```
                                                          Sau3A I                  Sau3A I
    Sau3A I                                               Mbo I                    Mbo I
    Mbo I        HgiA I                                   Dpn I                    Dpn I
    Dpn I        Bsp1286 I                   Taq I        BstY I                   BstY I
Mbo II           ApaL I            Mae III                Alw I        NspB II     Alw I
CTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC   2720
GACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGG
2643             2656              2668     2674          2681         2690        2699
    2646         2656                                     2681                     2698
    2646         2656                                      2682                    2699
    2646                                                   2682                    2699
                                                           2682                    2699

                                                                                     ScrF I
                                                                                     Nci I
                                                                                     Msp I
                                                                                     Hpa II
                                                         BstU I                      Bcn I
   Mae II                             Dra I              HinP I                 Hga I
  Xmn I                  HgiA I       Mse I              Hha I                  Aha I
Mbo II                   Bsp1286 I
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGA   2800
CTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCT
2721                     2741          2749              2767                   2787
  2724                   2741         2748               2767                   2787
    2726                                                  2768                     2791
                                                                                   2791
                                                                                   2791
                                                                                   2791
                                                                                   2791

                                                          Hph I
                                                      Rsa I
               Fnu4H I            Dde I              Sca I         Mae III      SfaN I    Fok I
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATG   2880
CGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTAC
               2812               2827               2845          2856         2867      2876
```

Fig. 7L

```
                                                        2848
                                                              2850
                                                                    Hae III                         Sau3A I
                                                                   Gdi I                            Mbo I
                         Fnu4H I                                   Eae I                            Dpn I
Nla III                  Bbv I         Nla III                    Fnu4H I                           Pvu I
|                        |             |                          |||                               ||
GCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC  2960
CGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAG
|                        |             |                          |||                               ||
2882                     2907          2918                       2934                              2956
                         2907                                      2935                              2957
                                                                   2935                              2957
                                                                    2936                             2957

                                         Sau3A I
                                         Mbo I
   Sau96 I                               Dpn I                  Sau3A I       Msp I
   Ava II                               Alw I   Mae III         Mbo I         Hpa II
Mnl I        Alu I                 Nla III  Nla III             Dpn I    Nla IV  Alu I
||           |                     |    ||  |  |                |        |   |   |
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCT  3040
CCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACCCTTGGCCTCGA
||           |                     |    ||  |  |                |        |   |   |
2962         2974                  2996     3006                3021     3029    3037
 2964                                   3002   3009             3021         3033
 2964                                    3003                   3021         3033
                                         3003
                                         3003

                                                                   HinP I
                                                                   Hha I
                                                                  Fsp I
                   Mae III      SfaN I                        Mae II             Mse I
                   |            |                             |  ||              |
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTG  3120
CTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGATAATTGAC
                   |            |                             |  ||              |
                   3067         3077                          3099               3114
                                                                 3103
                                                                  3104
                                                                  3104
```

Fig. 7M

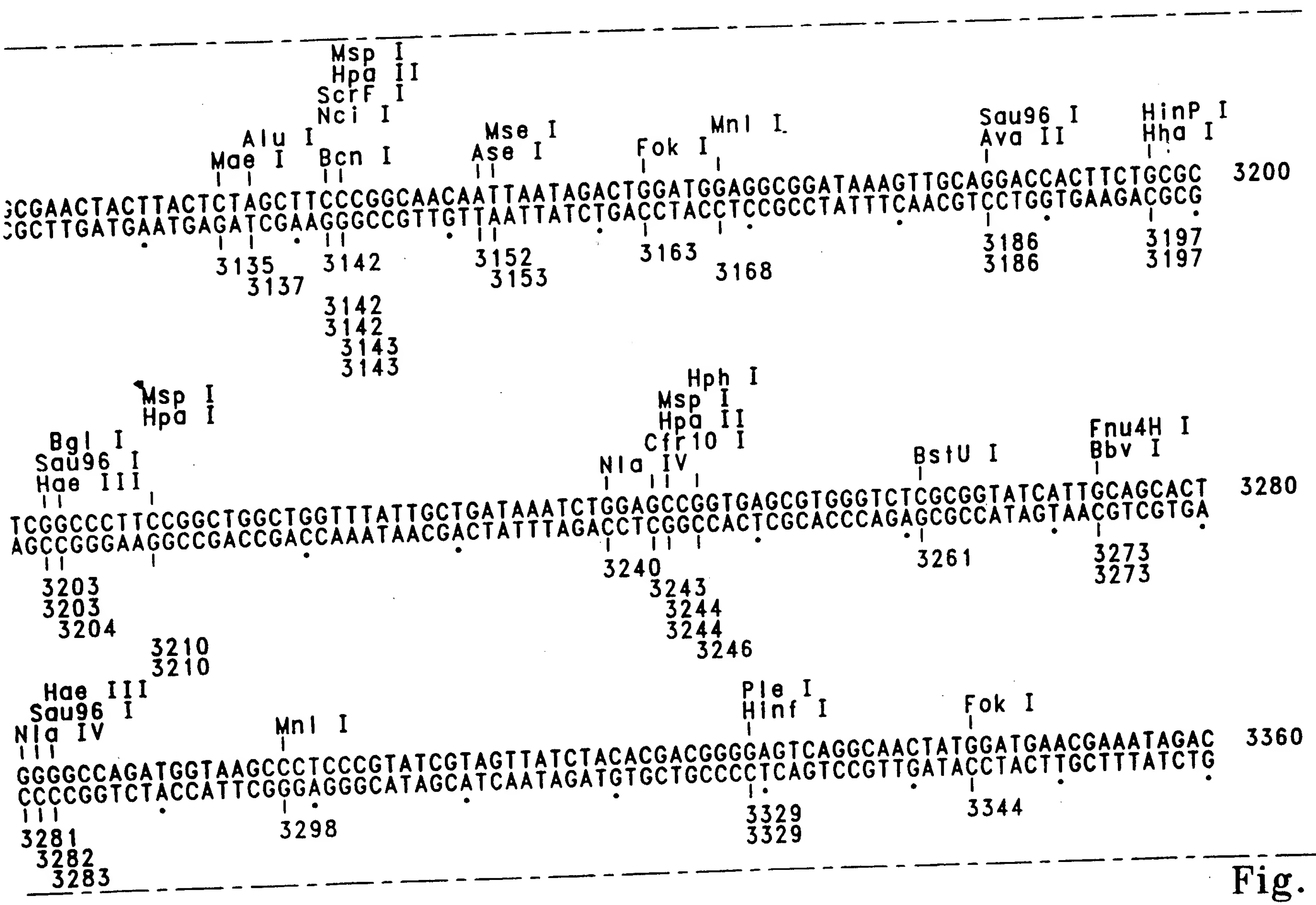

Msp I
Hpa II
ScrF I
Nci I
Alu I
Mae I
Bcn I
Mse I
Ase I
Fok I
Mnl I
Sau96 I
Ava II
HinP I
Hha I
GCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC 3200
CGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCG
3135
3137
3142
3142
3142
3143
3143
3152
3153
3163
3168
3186
3186
3197
3197
Msp I
Hpa I
Bgl I
Sau96 I
Hae III
Hph I
Msp I
Hpa II
Cfr10 I
Nla IV
BstU I
Fnu4H I
Bbv I
TCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACT 3280
AGCCGGGAAGGCCGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGA
3203
3203
3204
3210
3210
3240
3243
3244
3244
3246
3261
3273
3273
Hae III
Sau96 I
Nla IV
Mnl I
Ple I
Hinf I
Fok I
GGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGAC 3360
CCCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTG
3281
3282
3283
3298
3329
3329
3344

Fig. 7N

```
            Dde I
Sau3A I
Mbo I              Mnl I
Dpn I       Nla IV
            Ban I          Mse I      Mae III                               Dra I
AGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGAT 3440
TCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTA
3362        3375           3388       3398                                  3440
3362        3375
3362           3379
    3367

                                          Sau3A I
                              Mae I       Mbo I
                            Sau3A I       Dpn I
                            Mbo I         Alw I
                            Dpn I         BstY I
                    Mse I   BstY I      Mbo II
Mse I              Dra I    Alw I   Hph I                     Nla III               Mae II
            Mse I                                             BspH I            Mse I
TTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA 3520
AATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACT
3441        3455            3466    3473                      3496              3512
                3459        3466      3476                      3497              3515
                  3460        3467      3478
                              3467        3479
                              3467        3479
                                3470      3479
                                          3479

                                                 Sau3A I
                                                 Mbo I
                                    Sau3A I      Dpn I         Sau3A I
                                    Mbo I          Mbo II      Mbo I
              Hga I                 Dpn I        BstY I        Dpn I          BstU I
          Dde I                                  Alw I         Alw I          HinP I
                                                               BstY I         Hha I
GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT 3600
CAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGA
```

Fig. 70

```
3533 3537 3557 3557 3557 3564 3564 3567 3565 3565 3565 3575 3576 3576 3576 3576 3590 3590 3591

                                                          Sau3A I
                                                          Mbo I
                                                          Dpn I
                                                        Alw I
                                                      Msp I
 Tth111 II                                            Hpa II
Fnu4H I                                  Tth111 II                   Alu I
Bby I    Tth111 II           NspB II
GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG   3680
CGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTC
3601     3609                3631        3642        3648            3658
3601                                                 3648
 3603                                                 3650
                                                       3651
                                                       3651
                                                       3651

                                                                Hae III
                   HinP I                     Mae I           Hae I
Mae III            Hha I
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC   3760
CATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAG
3681               3699                       3723            3740
                   3699                                        3741

                                                                        ScrF I
                                                                        Nci I
                                                  Fnu4H I               Msp I
                                              Fnu4H I                   Hpa I
                                          AlwN I  Bbv I                 Bcn I
                   Mnl I              Mae III  Bbv I
TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCG   3840
ACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGC
```

Fig. 7P

```
  3779        3797  3802  3807  3810                        3838
                          3807  3810                        3838
                                                            3838
                                                            3838
                                                            3838

                                NspB II
                               Fnu4H I
                Msp I          Bbv I                 HgiA I
 Ple I          Hpa II        HinP I                 Bsp1286 I           Alu I
 Hinf I      Mae III          Hha I                  ApaL I
GGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTG  3920
CCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAAC
 3846        3860  3864        3873                      3902               3915
 3846              3864        3873                      3902
                                3875                     3902
                                3875
                                 3876

                                              HinP I
                                              Hha I
             Dde I                            Hae II
GAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGC  4000
CTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGTAACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCG
             3942                             3972
                                               3973
                                               3973
                                                 Sec I
                                                 ScrF I
                                  Mnl I          EcoR II     ScrF I
      Msp I  Fnu4H I         HinP I              BstN I      EcoR II
      Hpa II                 Hha I          Alu I            BstN I
GGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTT  4080
CCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAA
      4011                   4040           4051             4069
```

Fig. 7Q

```
4011 4018 4040 4046 4056 4056 4056 4056 4069 4069

        Mnl I    Hga I Taq I    SfaN I                 Nla IV
ATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAA 4160
TATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTT
        4103     4115 4118         4129                  4147

                              Hae III
                              Hae I
                              ScrF I                            Nla III
    Hae III                   EcoR II                           NspH I
    Fnu4H I                   BstN I          Hae III           Nsp7524 I
    BstU I          Nla IV                    Hae I             Afl III
AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCC 4240
TTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGG
    4172 4173 4175  4186  4190 4190 4190 4192 4193  4203 4204  4216 4216 4216 4217

                                                                        Ple I
                                                                        Fnu4H I
                                                  Fnu4H I               Bbv I
                                                  Bbv I                 HinP I
Hinf I                           Alu I            Fnu4H I               Hha I   Hinf I
TGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT 4320
ACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCA
4242                             4277             4291 4294 4294    4310 4310 4312  4317
```

Fig. 7R

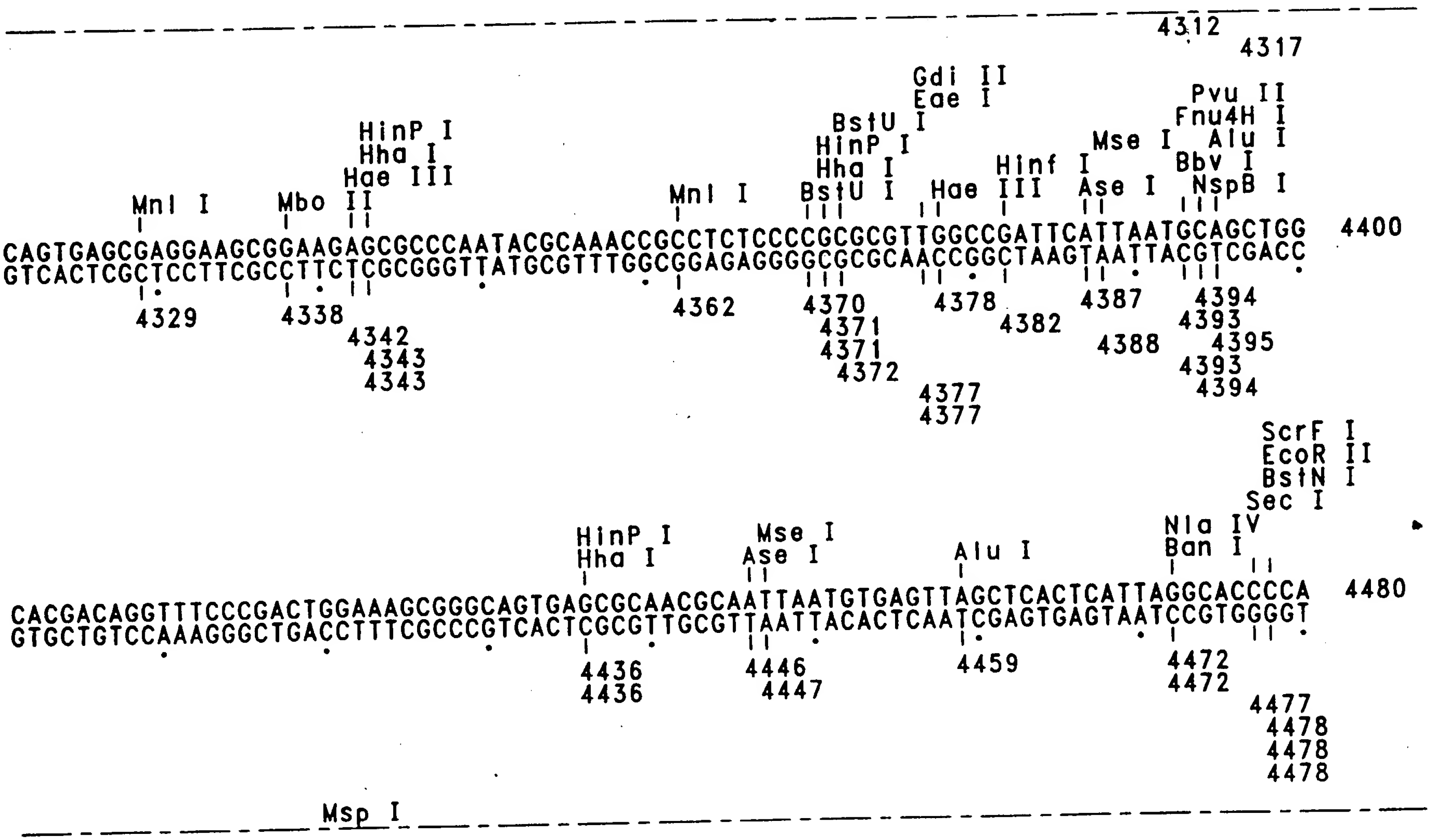
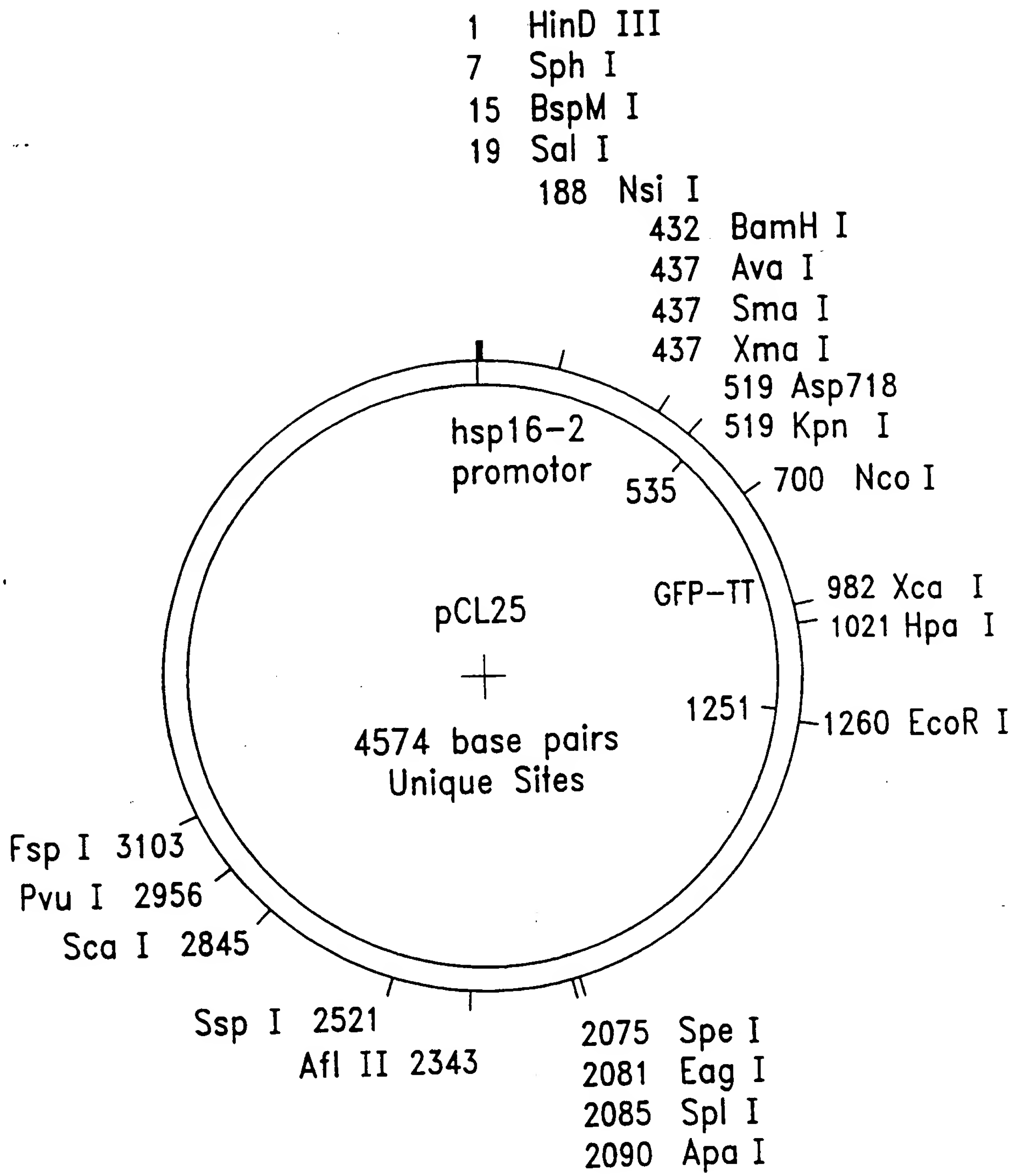

4312
4317
Mnl I
Mbo II
HinP I
Hha I
Hae III
Mnl I
BstU I
HinP I
Hha I
BstU I
Gdi II
Eae I
Hae III
Hinf I
Mse I
Ase I
Pvu II
Fnu4H I
Alu I
Bbv I
NspB I
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGG
GTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACC
4400
4329
4338
4342
4343
4343
4362
4370
4371
4371
4372
4377
4377
4378
4382
4387
4388
4394
4393
4395
4393
4394
HinP I
Hha I
Mse I
Ase I
Alu I
Nla IV
Ban I
ScrF I
EcoR II
BstN I
Sec I
CACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCA
GTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGT
4480
4436
4436
4446
4447
4459
4472
4472
4477
4478
4478
4478
Msp I

Fig. 7S

```
                    Hpa II                                                               Alu I
GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATG   4560
CCGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATAC
                    4500                                                                 4554
                    4500

  Nla III
ACCATGATTACGCC   4574
TGGTACTAATGCGG
  4563
```

Fig. 7T



1 HinD III
7 Sph I
15 BspM I
19 Sal I
188 Nsi I
432 BamH I
437 Ava I
437 Sma I
437 Xma I
519 Asp718
519 Kpn I
700 Nco I
982 Xca I
1021 Hpa I
1260 EcoR I
hsp16-2
promotor
535
GFP-TT
1251
pCL25
4574 base pairs
Unique Sites
Fsp I 3103
Pvu I 2956
Sca I 2845
Ssp I 2521
Afl II 2343
2075 Spe I
2081 Eag I
2085 Spl I
2090 Apa I

Fig. 8

```
   1 atcacaatga catccttatg tatatgcatt ctttttattg tttcctgata tttattctct
  61 ctttatcatt tcactttttc ttggcactcg tccaaaaatc tcaattaaaa atttctggat
 121 atatctttag atctaactga aaatttccag atgaccctaa ctacggcgac gtccggcgcc
 181 attgtatttt ctggagccac tttgttggtt tctcttttg ccgctgcttc gctttacagt
 241 caagtttcca atatctggaa tgagctggat gccgaaatcg ccaacttcag agtgagtttc
 301 ctatatagca tttctaattc actgataatt tcaattattt agagtctcac tgaagacatg
 361 tgggttgata tggttaaact tggagcagga accgcttcca accgtgtgag acgtcaacaa
 421 tatggaggat atggagccac tggtgttcag ccaccagcac caactccaaa cccatatgga
 481 ggatatggag caagccagcc agctccacca gagaaattcc cagatggtat accaaatgga
 541 ggaaatcaac caaagttccc aggaggtgga ttcccagatg gtccattccc gaatggagga
 601 ggaccacgtg gaggaaatca atgtcaatgc actgttgaga actcatgccc accaggacca
 661 gctggaccag aaggagagga aggaccagat gggcacgatg gacaagacgg agtcccagga
 721 tttgacggaa aagatgccga agatgttcaa aacacccac caacaggatg cttcacctgt
 781 ccacaaggac cacttggacc acaaggacca aatggagctc caggactcag aggaatgcgc
 841 ggagctcgtg gacaaccagg acgtccagga agagacggaa acccaggaat gccaggagac
 901 tgtggaccac caggggcccc aggatccgat ggaaagccag gatctccagg aggaaaggga
 961 gacgatggag agagaccatt gggccgccca ggaccaagag gaccaccagg agaggctggg
1021 ccagaaggac cacagggacc aactggaaga gatgcttatc caggacagtc tggaccacaa
1081 ggagagccag gccttcaagg atatggagga gctgctggag aggacgggta ggatgaaaaa
1141 aatattaaac gacttcttgt aaatataaaa atttcagacc agaaggacca ccaggagccc
1201 caggacttcc aggaaaagat gccgaatact gcaaatgccc aggaagagaa ggagatgctg
1261 gacgaagtgc cagacgtcat cgcaaattcc aattgtagac aaattcatga catttttcca
1321 aaaaataaaa catacttctc aaaatttttt gttttgtatt tggtttacca tggatgttaa
1381 gaactttctt ggggaaaata attaaaatac aaaactgtat a
```

Fig. 9

MTLTTATSGAIVFSGATLLVSLFAAASLYSQVSNIWNELDAEIANFRSLTEDMWVDMVK
LGAGTASNRVRRQQYGGYGATGVQPPAPTPNPYGGYGASQPAPPEKFPDGIPNGGNQP
KFPGGGFPDGPFPNGGGPRGGNQCQCTVENSCPPGPAGPEGEEGPDGHDGQDGVPGFD
GKDAEDVQNTPPTGCFTCPQGPLGPQGPNGAPGLRGMRGARGQPGRPGRDGNPGMPG
DCGPPGAPGSDGKPGSPGGKGDDGERPLGRPGPRGPPGEAGPEGPQGPTGRDAYPGQSG
PQGEPGLQGYGGAAGEDGPEGPPGAPGLPGKDAEYCKCPGREGDAGRSARRHRKFQL

Fig. 10

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00, A01K 67/033, C07K 14/435, 14/47

A3

(11) International Publication Number: WO 98/28971

(43) International Publication Date: 9 July 1998 (09.07.98)

(21) International Application Number: PCT/US97/23819

(22) International Filing Date: 31 December 1997 (31.12.97)

(30) Priority Data:

08/778,645	3 January 1997 (03.01.97)	US
60/080,222	3 January 1997 (03.01.97)	US

(71) Applicant (*for all designated States except US*): UNIVERSITY TECHNOLOGY CORPORATION [US/US]; Suite 250, 3101 Iris Avenue, Boulder, CO 80301 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): LINK, Christopher [US/US]; 1837 Spruce Street, Boulder, CO 80302 (US).

(74) Agents: MONROY, Gladys, H. et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304–1018 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

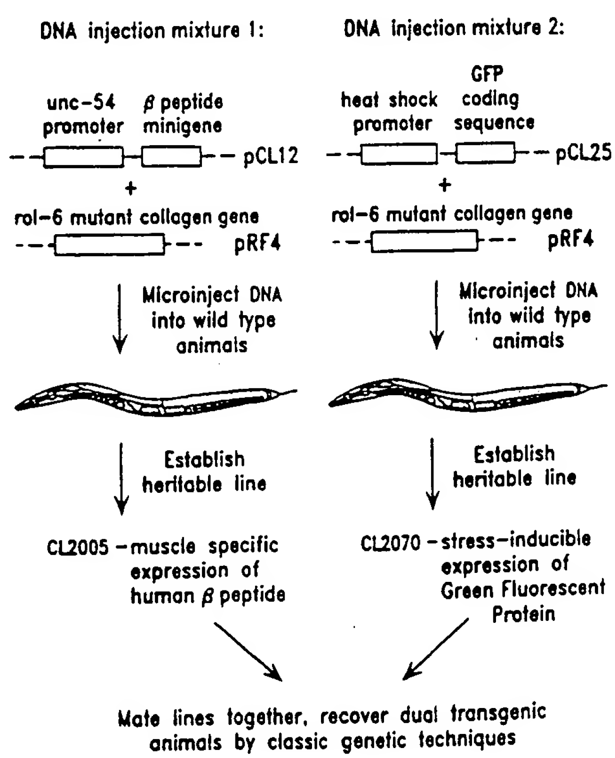
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 29 October 1998 (29.10.98)

(54) Title: TRANSGENIC ANIMALS FOR *IN VIVO* VISUALIZATION OF BETA-AMYLOID TOXICITY

(57) Abstract

The present invention relates to methods and compositions for visualization of the toxic effects of transgenes *in vivo*. In particular, the present invention provides methods and compositions for the production and use of transgenic, including dually transgenic, *Caenorhabditis elegans* for visualization of the toxic effects of β-amyloid accumulation *in vivo*.



INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/23819

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/00 A01K67/033 C07K14/435 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A01K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CA 2 088 379 A (CANDIDO EDWARD P M ;STRINGHAM EVE G (CA); JONES DONALD (CA)) 30 July 1994	1-4,6,7, 13-15,20
Y	see the whole document	1-23

Y	LINK, C.D.: "Expression of human beta-amyloid peptide in transgenic Caenorhabditis elegans" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 92, no. 20, 26 September 1995, pages 9368-9372, XP002065639 WASHINGTON US cited in the application	1-8, 10-15, 18-20, 22,23
X	see the whole document	11-15

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 September 1998	11 09 1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Chambonnet, F

4

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No.
Y	LINK, C.D.: "Transgenic Caenorhabditis elegans as model system to study amyloid formation and toxicity" NEUROBIOLOGY OF AGING, vol. 15, no. suppl. 1, 29 July 1994, page S33 XP002065640 see the whole document ---	1,5-7
X	WO 95 11968 A (ATHENA NEUROSCIENCES INC ;LILLY CO ELI (US)) 4 May 1995 see page 38, line 3 - page 39, line 22 ---	11,12,19
Y	WO 93 23533 A (TSI CORP) 25 November 1993 see page 27, paragraph 3 - page 28, paragraph 2 ---	11-15
Y	CHALFIE M ET AL: "GREEN FLUORESCENT PROTEIN AS A MARKER FOR GENE EXPRESSION" SCIENCE, vol. 263, no. 5148, 11 February 1994, pages 802-805, XP002003599 see the whole document ---	1,2,6,7, 9,11, 13-17, 20,21
Y	MOSS, J. B. ET AL.: "Green fluorescent protein marks skeletal muscle in murine cell lines and zebrafish" GENE, vol. 173, no. 1, 1996, pages 89-98, XP002065641 AMSTERDAM NL see the whole document -----	1,2,9, 13,16,17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 97/23819

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-10 and partially 11-23

a transparent animal comprising :
(a) a toxic transgene;
(b) an inducible promotor operably linked to a reporter gene; wherein the toxic transgene induces the promotor and wherein the expression of the reporter gene is detectable in vivo; and
a method for providing such a dually transgenic transparent non-human animal comprising :

a) providing :
i) a first and a second non-human transparent animals;
ii) a first toxic transgene; and
iii) a second transgene, comprising an inducible promotor operably linked to a reporter gene, wherein the toxic transgene induces the promotor and wherein the expression of the reporter gene is detectable in vivo,

b) introducing said first toxic transgene into the genome of said first non-human animal to produce a first transgenic animal and introducing said second transgene into the genome of said second non-human animal to produce a second transgenic animal; and

c) mating said first transgenic animal with said second transgenic animal to produce a dually transgenic transparent non-human animal wherein said toxic transgene is expressed and induces the promotor so that the said reporter gene is also expressed and detectable in vivo;

2. Claims: partially 11-23

a method to produce a dually transgenic non-human animal with a toxic transgene and a reporter are expressed comprising :

a) providing :
i) a first and a second non-human animal;
ii) a first toxic transgene; and
iii) a second transgene, comprising a reporter,

b) introducing said first toxic transgene into the genome of said first non-human animal to produce a first transgenic animal and introducing said second transgene into the genome of said second non-human animal to produce a second transgenic animal; and

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

c) mating said first transgenic animal with said second transgenic animal to produce a dually transgenic non-human animal wherein said toxic transgene and said reporter are expressed

in the cases non relevant of the first described subject, that means where it is not a transparent animal comprising :

(a) a toxic transgene;
(b) an inducible promotor operably linked to a reporter gene; wherein the toxic transgene induces the promotor and wherein the expression of the reporter gene is detectable in vivo.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/23819

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
CA 2088379	A	30-07-1994	NONE			
WO 9511968	A	04-05-1995	US	5604102	A	18-02-1997
			US	5612486	A	18-03-1997
			AU	8079894	A	22-05-1995
			AU	8080994	A	22-05-1995
			CA	2174429	A	04-05-1995
			CA	2174632	A	04-05-1995
			EP	0736106	A	09-10-1996
			EP	0730643	A	11-09-1996
			JP	9508196	T	19-08-1997
			JP	9507746	T	12-08-1997
			WO	9511994	A	04-05-1995
WO 9323533	A	25-11-1993	AU	4240293	A	13-12-1993